







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 31, 2004, 04:45:19 ; Search time 4126 Seconds  
(without alignments)  
4222.952 Million cell updates/sec

Title: US-10-087-080-32

Perfect score: 2123

Sequence: 1 MKLEVFVPRAAHGDKQSDL.....AALVRRPGPHLSYPVETLLA 402

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlsp  
-O=/cgn2.1/USPTO\_spool\_p/US10087080/runat\_29032004\_114946\_24387/app\_query.fasta\_1.583  
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087080@cgn2.1\_4988@runat\_29032004\_114946\_24387 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2123	100.0	1448	9	AF153341 Homo sapi
2	2093.5	98.6	62761	9	AL499606 Human DNA
3	2081.5	98.0	2322	9	BC053850 Homo sapi
4	2081.5	98.0	4109	9	AF225950 Homo sapi
5	1597	75.2	1513	10	AF154426 Mus muscu
6	1597	75.2	2046	10	BC047155 Mus muscu
7	1597	75.2	187709	10	AL589738 Mouse DNA
8	1589	74.8	4763	10	AF010405 Mus muscu
9	1534	72.3	1576	10	AF153193 Rattus ra
10	1534	72.3	195184	2	AC119497 Rattus no
11	1412.5	66.5	1760	10	L13201 Rattus norv
12	1372	64.6	182152	2	AC067929 Homo sapi
13	1159	54.6	178168	2	AC012350 Homo sapi
14	956.5	45.1	74969	2	AC016269 Homo sapi
15	685.5	32.3	62848	5	BX510653 Zebrafish
16	561.5	26.4	2222	3	AK112352 Clona int
17	522	24.6	20236	5	GU047276 Gallus gall
18	494.5	23.3	106239	9	AL607122 Human DNA
19	494.5	23.3	153604	2	AC022754 Homo sapi
20	493.5	23.2	4259	9	AF042832 Homo sapi
21	488.5	23.0	2272	9	HSU59832 Human trans
22	488.5	23.0	5181	6	AX336763 Sequence
23	487.5	23.0	5181	9	HSU59831 Human trans
24	487.5	23.0	1593	10	AF023915 Mus muscu
25	487.5	23.0	201142	10	AL670035 Mouse DNA
26	486	22.9	1708	5	GU372722 Gallus gall
27	484.5	22.8	2437	10	MUSBF2TF
28	479.5	22.6	173073	9	AC099522 Homo sapi
29	479.5	22.6	182835	10	AC121839 Mus muscu
30	473.5	22.3	7005	10	AF142647
31	468.5	22.1	3946	6	AR102796 Sequence
32	468.5	22.1	3946	6	AR140209 Sequence
33	468.5	22.1	3946	6	BD226150 Glaucoma
34	468.5	22.1	3946	9	AF048693 Homo sapi
35	468.5	22.1	104729	9	HS118818 Human DNA
36	467.5	22.0	3582	9	HSFREAC1
37	467.5	22.0	120429	9	HS668J24
38	467	22.0	1465	5	GGU37274
39	466.5	22.0	2187	6	AX578019 Sequence
40	466.5	22.0	2187	9	HSU13220 Homo sapien
41	465.5	21.9	209010	2	AC123435 Rattus no
42	462.5	21.8	1673	9	HS11F2
43	462	21.8	2011	9	AF275722 Homo sapien
44	462	21.8	142728	9	HSU792G4 Human DNA
45	460	21.7	217762	10	AL806523 Mouse DNA

ALIGNMENTS

RESULT 1

AF153341 1448 bp DNA linear PRI 24-AUG-2001  
LOCUS Homo sapiens winged helix/forkhead transcription factor (HFH1)  
DEFINITION gene, complete cds.  
ACCESSION AF153341  
VERSION AF153341.1 GI:8499092  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1448)  
Justice M.J. and Chakravarti, A.  
TITLE The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice  
JOURNAL Genesis 29 (4), 163-171 (2001)  
MEDLINE 21207067  
PUBMED 11309849  
REFERENCE  
AUTHORS 2 (bases 1 to 1448)  
Hong H.-K., Noveroske, J.K., Justice, M.J. and Chakravarti, A.  
TITLE Direct Submission  
JOURNAL Submitted (21-MAY-1999) Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4955, USA  
FEATURES  
source  
1..1448  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="6"  
/map="6p25"  
<233..>1441  
/gene="HFH1"  
<233..>1441  
/gene="HFH1"  
/product="winged helix/forkhead transcription factor"  
233..1441  
/gene="HFH1"  
/codon\_start=1  
/product="winged helix/forkhead transcription factor"  
/db\_xref="GI:8499092"  
/translation="MKLEVFVPAHGDQKQSDLEGAGSDAPSLAAGDDSLGSDG  
DCAAKPSAGGARDTQDGEAGGPGAEAEI PAARAAAVVAEGAEAGAAGPGAGGA  
GSGEGARSPYTRRPKPYSYIALIAMAIRDSAGRLTLAEINVMGKFFPRGSYT  
GWNSVRHLSLNDKVLKRDPSRPWMDNVMNPNSEYTFADGVFRRRKRLSHR  
APVAPGLSPEEPAGLPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP  
RKPFRRRLRDTPGTLGWAAPCPPLPAFPALLPAAPCPPLPAFPALLPAAPCPPL  
AREAEPVTPAPLLAPLPAAPAPKPLRPAAGGAHLICPLRLPAALQALVRRPPEH  
LSYPVETLLA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 3 53e-50 Length: 1448  
Score: 2123.00 Matches: 402  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-10-087-080-32 (1-402) x AF153341 (1-1448)  
Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
Db 233 ATGAAGTTGGAGGTGTTCTGTCCTCGCGCGCCACAGGGGACAGCAGGCGCAGTGACCTG 292  
Qy 21 GluGlyAlaGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40  
Db 293 GAGGCGCGGCGGCGACGACGCGCGCTCCCGCTGTGCGCGCGGAGACGACTCCCTG 352  
Qy 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGln 60  
Db 353 GGCTCAGATGGGACTCGCGGCGCAAGCGCTCCGCGGCGCGCGCGCGCGCGCGCGCG 412

61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyValaGluGluAlaIleProAlaAla 80  
Db 413 GGGACGGCGCAACAGAGTGGCGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 472  
Qy 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100  
Db 473 GCTGCTGCACGGTGGTGGCGAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 532  
Qy 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120  
Db 533 GCGCGCGGAGCGCGGAGGGTGCACGACCAAGCCATATACGCGCGCGCGCGCGCGCGCG 592  
Qy 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140  
Db 593 TACTCGTACATCGCTCATCGCATGGCCCAACCTTTCGCTCAACGACTGCTTCGTCAGG 652  
Qy 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrThr 160  
Db 653 CTGGCGGAGATCAACGAGTACCTCATGGGCAAGTCCCTTTTCCGCGCGCAGCTACAG 712  
Qy 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180  
Db 713 GGCTGGCGCAACTCCGTCGCGCCCAACCTTTCGCTCAACGACTGCTTCGTCAGGTCGT 772  
Qy 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlu 200  
Db 773 CGGACCCCTCGCGCGCCCTGGGCGAGGCAAGCACTACTGGATGCTCAACCCACAGCGAG 832  
Qy 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysArgLysArgLysArgAlaPro 220  
Db 833 TACACCTTCGCGAGCGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 892  
Qy 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240  
Db 893 GTCCCCCGCGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 952  
Qy 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluAlaArgAla 260  
Db 953 GCGCGCGCGCGCGCGCTTCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1012  
Qy 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280  
Db 1013 AGCCCCCGGGCAAGTTCTCCAGCTCTTCGCCATCGACAGCATCTTCGCGAGCCCTTC 1072  
Qy 281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro 300  
Db 1073 GCGAGCGCTCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1132  
Qy 301 CysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeu 320  
Db 1133 TGCCGCGCGCTGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1192  
Qy 321 ProLeuCysAlaTyrGlyAlaGlyValProAlaArgLeuGlyAlaArgGluAlaGluVal 340  
Db 1193 CCGCTCTGCGGCTACG 1252  
Qy 341 ProProThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaProAlaLysPro 360  
Db 1253 CCACCGACCGCGCGCGCGCTCTCTGCTTGCACCTCTCCCGCGCGCGCGCGCGCGCGCG 1312  
Qy 361 LeuArgGlyProAlaAlaGlyValaHisLeuTyrCysProLeuArgLeuProAlaAla 380  
Db 1313 CTCGAGGCG 1372  
Qy 381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThrLeu 400  
Db 1373 CTGACGCGCGCTTAGTCCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1432  
Qy 401 LeuAla 402  
Db 1433 CTAGCT 1438

```

RESULT 2
LOCUS      AL499606                62761 bp    DNA        linear        PRI 19-DEC-2001
DEFINITION Human DNA sequence from clone RP11-13J16 on chromosome 6, complete
sequence.
ACCESSION  AL499606
KEYWORDS   HTG
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 62761)
AUTHORS    Corby, N.
TITLE      Direct Submission
JOURNAL    Submitted (19-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Dec 20, 2001 this sequence version replaced gi:17902904.
COMMENT    During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em, EMBL; Sw,
            SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-13J16 is from the library RPC1-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBAC3.6
FEATURES             source
     source           1..62761
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-13J16"
                     /clone_lib="RPC1-11.1"
                     /1711..17234
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature     17634..17641
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature     21165..21540
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Assembly consistent with restriction digest data."
     misc_feature     21466
                     /note="Tandem repeat. Forced join. Gap size estimated to

```

```

be approximately 430bp by restriction digest data."
/Note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
60092
/Note="Tandem repeat. Forced join. Gap size estimated to
be approximately 120bp by restriction digest data."

ORIGIN
Alignment Scores:
Pred. No.:      4,1e-48      Length:      62761
Score:          2093.50      Matches:    398
Percent Similarity: 99.01%    Conservative: 1
Best Local Similarity: 98.76%  Mismatches:   3
Query Match:    98.61%      Indels:     1
DB:             9           Gaps:       1

US-10-087-080-32 (1-402) x AL499606 (1-62761)
QY      1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db      16945 ATGAGCTTGGAGGTGTTCCTCGCGCGGCCCAAGCAGGCGAGTACCTG 17004.
QY      21 GluGlyAlaGlyCysAspAlaProSerProLeuSerAlaAlaGlyAspAspSer 40
Db      17005 GAGGGCGCGGGCGGACCGCGCGTCCCGCGTGTGGCGGGGAGGAGGACTCCCTG 17064
QY      41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr 59
Db      17065 GGTCTAGATGGGACTTCGCGGCCCAACACCGCGCGCGGGCGCGCCAGAGATACG 17124
QY      60 GlnGlyAspGlyClnGlnSerAlaGlyGlyGlyProGlyAlaGluAlaLalleProAla 79
Db      17125 CAGGGCGAGCGGCAACAGAGTGGCGGAGCGCGCGCGGAGGAGGCGATCCCGGCA 17184
QY      80 AlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAla 99
Db      17185 GCAGCTCTGCGAGCGGTGTGGCGGAGGCGCGGAGGCGCGGGCGCGGCGCGCG 17244
QY      100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysPro 119
Db      17245 GCGCGCGCGGGAGCGCGGAGGTGCACGAGCAAGCCATATACGCGGGCGGCGCAAGCCC 17304
QY      120 ProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeu 139
Db      17305 CCTACTCGTACATCGCTCATCGCATCGGCGCATCCGCGACTCGCGCGGGCGCGCTTG 17364
QY      140 ThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPhePheProPhePheArgGlySerTyr 159
Db      17365 ACCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTTCCGCGGAGCTAC 17424
QY      160 ThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysVal 179
Db      17425 ACGGGCTGGCGCAACTCCGTGGCGCCACACCTTTCCTCAACGACTGCTTCGCAAGGTG 17484
QY      180 LeuArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSer 199
Db      17485 CTCGGCGACCCCTCGCGGCGCTGGGGCAAGGCAACAATACTGGATGCTCAACCCCAACAGC 17544
QY      200 GluTyrThrPheAlaAspGlyValPheArgArgArgArgArgLysSerHisArgAla 219
Db      17545 GAGTACACCTTCGCCGACGCGGTCTTCCGCGCGCGCGCAAGCGCTCAGCACCGCGCG 17604
QY      220 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProPro 239
Db      17605 CCGGTCCCGCGCGCGCGGTGTGGGCCCGGAGGCGCGCGGCGCTCCCGCGCGCGCGCG 17664
QY      240 ProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 259
Db      17665 CCGCGCGCGCGCGCGCGCTCGCGCGCGCATGCGCTCGCGCGCGCGCGCGGAGGAGCGC 17724
QY      260 AlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArgLysPro 279

```

Db	17725	GCAGCGCCGCGGCAAGTTCTCCAGCTCTTCGCGCATCGACAGCATCTCGGCAAGCCC	17784
Qy	280	PheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAla	299
Db	17785	TTCCGAGCGCGCGCTCAGGACACGGCCCGGAGCACTTCAGTGGGCGCGCGC	17844
Qy	300	ProCysProLeuProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeu	319
Db	17845	CCCTGCGCGCGCTGCGCGCTTCCTCCCGCGCTCTCCCGCGCGCGCTCGAGGCGCTG	17904
Qy	320	LeuProLeuCysAlaTyArgAlaGlyGluProAlaAlaArgLeuGlyAlaAargGluAlaGlu	339
Db	17905	CTCGCGCTCTGCGGTACGGCGGCGGAGCGCGCGGCTGGCGCGCGGAGCGCGAG	17964
Qy	340	ValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLys	359
Db	17965	GTGCCACCGACCGCGCGCTCTCTGCTGTGACCTCTCCCGCGCGCGCGCCCGCCCAAG	18024
Qy	360	ProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyCysProLeuArgLeuProAla	379
Db	18025	CCACTCGAGGCGCGCGCGCGCGCGGCGGACCTGTACTGCCCTCGCGCTGCGCGCA	18084
Qy	380	AlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyProValGluThr	399
Db	18085	GCCTCGAGCGCGCTCAGTCCGCGCGCTGGCGCGCACTGCCGTACCGGTGGAGACG	18144
Qy	400	LeuLeuAla 402	
Db	18145	CTCCTAGCC 18153	
RESULT 3			
BC053850			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			

```
QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerIleu 40
Db 326 GAGGCGCGGCGGCGAGCGCGCGGCTCCCGCTGTGCGCGGCGGAGACGACTCCCTG 385
QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyAlaArgAspThr 59
Db 386 GGCTCAGATGGGACTGCGCGGCAACAGCGCGCGCGCGCGCGCGCGCGGATCCG 445
QY 60 GlnGlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGluAlaIleProAla 79
Db 446 CCGGCGCGGCGGCGGAGTGCAGGCGGCGGCGGCGGCGGCGGAGGAGCGATCCCGGCA 505
QY 80 AlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAla 99
Db 506 GCAGCTGTCTGACGGTGTGCGGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCG 565
QY 100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysPro 119
Db 566 GCGCGCGCGGCGGAGCGGCGGAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 625
QY 120 ProTyrSerTyrIleAlaLeuAlaMetAlaIleArgAspSerAlaGlyGlyArgLeu 139
Db 626 CCTTACTGTTACATCGGCTCATGCGCATGCGGCTCGGCGGCGGCGGCGGCTTG 685
QY 140 ThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyr 159
Db 686 ACGCTGGCGGAGATCAACAGTACTCATGCGCAAGTTCGCTTTTCCGCGCGAGCTAC 745
QY 160 ThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValIleVal 179
Db 746 ACGGCTGGCGCAACTCCGTCGCGCAACACCTTTCGCTCAACGACTGCTCGTCAAGGTG 805
QY 180 LeuArgAspProSerArgTyrProGlyLysAspAsnTyrTrpMetLeuAsnProAsnSer 199
Db 806 CTGCGGAGACCTCTCGGCGCTGGGCGGAGCAACTACTGATGCTCAACCCCAACAGC 865
QY 200 GluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAla 219
Db 866 GAGTACACCTTTCGCGAGCGGGTCTTCGCGCGCGCGCGCAAGCGCTCCAGCCACCGCG 925
QY 220 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProPro 239
Db 926 CCGGTCCCGCGCGCGGCTCGGCGCGGAGAGCGCGCGGCTCCCGCGCGCGCGCG 985
QY 240 ProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 259
Db 986 CCGCGCGCGCGCGCGCGCTCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 1045
QY 260 AlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysPro 279
Db 1046 GCCAGCGCGCGGCAAGTTCTCAGCTCTCTCGCATCGACGATCTCTCGCGCAAGCCC 1105
QY 280 PheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTyrGlyAlaAla 299
Db 1106 TTCGCGAGCGCGCTCAGGACACGCGCGCGGAGCGCGCTTCACTGCGGCGCGCGCG 1165
QY 300 ProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeu 319
Db 1166 CCGTCCCGCGCGCTCGCGCGCTTCCCGCGCTCTCCCGCGCGCGCGCTCGAGCGCGCT 1225
QY 320 LeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGlu 339
Db 1226 CTGCGGCTCTCGGCTACGCGCGCGGAGCGCGCGGCTGGCGCGCGCGGCGCGAG 1285
QY 340 ValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaProAlaLys 359
Db 1286 GTGCCACCGAGCGCGCGCTCTCTGCTGCACTCTCCCGCGCGCGCGCGCGCGCGCAAG 1345
QY 360 ProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAla 379
Db 1346 CCATCCGAGCGCGCGCGCGCGGCGCGGAGCACTGTACTGCCCTCTCGCGCTGCGCGCA 1405
QY 380 AlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThr 399
```

```
Db 1406 GCCCTGACGGCGCTAGTCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 1465
QY 400 LeuLeuAla 402
Db 1466 CTCCTAGCC 1474

RESULT 4
AF225950 4109 bp DNA linear PRI 06-DEC-2001
LOCUS Hmo sapiens HNF-3/forkhead-like protein 1 (HFH1) gene, complete
DEFINITION cds.
ACCESSION AF225950
VERSION AF225950.1 GI:12655883
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4109)
AUTHORS Bieller,A., Pasche,B., Frank,S., Glaser,B., Kunz,J., Witt,K. and Zöll,B.
TITLE Isolation and Characterization of the Human Forkhead Gene FOXQ1
JOURNAL DNA Cell Biol. 20 (9), 555-561 (2001)
PUBMED 11747606
REFERENCE 2 (bases 1 to 4109)
AUTHORS Bieller,A., Pasche,B. and Zöll,B.
TITLE Direct Submission
JOURNAL Submitted (17-JAN-2000) Institute of Human Genetics, University of Goettingen, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany
FEATURES
source
1..4109
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p23-25"
<1737..>22948
/feature="HFH1"
<1737..>22948
/feature="HFH1"
/product="HNF-3/forkhead-like protein 1"
1737..2948
/feature="HFH1"
/feature="HFH1"
/feature="forkhead transcription factor; similar to Mus musculus HNF-3/forkhead homolog 1 like encoded by GenBank Accession Number AF010405 and Rattus norvegicus HNF-3/forkhead homolog-1 encoded by GenBank Accession Number L13201"
/codon_start=1
/product="HNF-3/forkhead-like protein 1"
/db_xref="GI:12655884"
/translation="MKLEVFVPRAAHGDQKGSGLLEGAGSDAPSLAAGDDSLGSDG
DCAANSFAGGAGARDPPGDSQAGGPGABEAIAPAAAAAVVAGAGAGAGGAGG
AGSGGARGKPYTRRPKPPSYIALIAMIARDSAGRLTLAEINYLAKGKFPFRGSGY
TCWRNSVRLNSLDCFKVLRDPSRPWKDNYMNLNSENVTADGVFRRRRKLISH
RAPVPAGIRPEERAEGLPAAPPAPAAAPSPMRSPARQERASAPACKESFATDSI
LRKPFRRLRDTAPGTLQWGAAPCPPLPAPFALLPAPCRALLPLCAYGAGEPARL
GAREAEVPTAPLLPAPAPAKPLRGPAAGAHLYCLRLPAPALQAASVRRPGL
HLPYEPVEILLA"
misc_feature
2058..2399
/feature="HFH1"
/feature="Region: DNA binding domain"

ORIGIN
Alignment Scores:
Pred. No.: 1,05e-48 Length: 4109
Score: 2081.50 Matches: 396
Percent Similarity: 98.51% Conservative: 1
Best Local Similarity: 98.26% Mismatches: 5
Query Match: 98.05% Indels: 1
DB: Gaps: 1
```

US-10-087-080-32 (1-402) x AF225950 (1-4109)	
QY	1 MetLeuLeuGluValPheValProArgAlaAlaHisGlyAspLeuGlnGlySerAspLeu 20
Db	1737 ATGAAGTTGGAGGTGTCTGCTCCTCGCGCGCCACGGGACACAGCAGGAGTGCACCTG 1796
QY	21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40
Db	1797 GAGGCGCGCGCGCAGCAGCAGCGCGCTCCCGCTGTCCGCGCGGAGAGCAGCTCCCTG 1856
QY	41 GlySerAspGlyAspCysAlaAla--LysProSerAlaGlyGlyGlyAlaArgAspThr 59
Db	1857 GGCTCAGATGGGACTCGCGCGCCACAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1916
QY	60 GlnGlyAspGlyGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIleProAla 79
Db	1917 CCGCGCGCGCGGAAACAGAGTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1976
QY	80 AlaAlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAla 99
Db	1977 GCAGCTGCTGACGCGGTGTGCGGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 2036
QY	100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysPro 119
Db	2037 GCGCGCGCGCGGAGCGCGCGGAGGTGCACGACAGCCATATACGCGCGCGCGCGCGCG 2096
QY	120 ProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyArgLeu 139
Db	2097 CCTACTCGTATACGCGCTCATCGCATGGCCATCGCGCATCGCGCGCGCGCGCGCGCTG 2156
QY	140 ThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyr 159
Db	2157 ACCTGGCGGAGNATCAACAGTACCTCATGGCGAGTTCCTCTTTTCGCGCGAGCTAC 2216
QY	160 ThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysVal 179
Db	2217 ACGGCTGGCGCAACTCCGTGGCGCACAACTTCGCTCAACGAGCTGCTTCGTCAGGTG 2276
QY	180 LeuArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSer 199
Db	2277 CTGCGCGACCCCTCGCGCGCCCTGGGCAAGGACAACTACTGATGCTCAACCCCAACAGC 2336
QY	200 GluTyrThrPheAlaAspGlyValPheArgArgArgArgLysArgLeuSerHisArgAla 219
Db	2337 GAGTACACCTTCGCGCAGCGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2396
QY	220 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProPro 239
Db	2397 CCGGTCGCGCGCGCGCGCTGCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2456
QY	240 ProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 259
Db	2457 CCGCGCGCGCGCGCGCGCGCTCGCCCGCATGCGCTCGCGCGCGCGCGCGCGCGCGCG 2516
QY	260 AlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysPro 279
Db	2517 GCCAGCGCGCGCGCGCAAGTCTCCAGCTCTTCGCCATCGACAGCATCTCGCGCAAGGCC 2576
QY	280 PheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTyrPglyAlaAla 299
Db	2577 TTCGCGAGCGCGCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2636
QY	300 ProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeu 319
Db	2637 CCTGCGCGCGCGCTGCGCGCGGTTCGCGCGGCTCCTCCCGCGCGCGCGCGCGCGCGCG 2696
QY	320 LeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGlu 339
Db	2697 CTGCGCGCTCTGCGCGTACGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2756
QY	340 ValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLys 359
Db	2757 GTCCACCGACCGCGCGCGCGCTCTGCTGTGACCTCTCCGCGCGCGCGCGCGCGCAAG 2816
QY	360 ProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAla 379
Db	2817 CCACTCCGAGGCG 2876
QY	380 AlaIeuGlnAlaAlaLeuValArgAspProGlyProHisLeuSerTyrProValGluThr 399
Db	2877 GCCTCCGAGCGCGCGCTCAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2936
QY	400 LeuLeuAla 402
Db	2937 CTCTAGCC 2945
RESULT 5	
AF154426	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
FEATURES	
source	
gene	
mRNA	
CDS	
ORIGIN	
Alignment Scores:	
Pred. No.:	
Score:	
Percent Similarity:	
Best Local Similarity:	
Query Match:	
DB:	
Length:	
Matches:	
Conservative:	
Mismatch:	
Indels:	
Gaps:	
1513 bp DNA linear ROD 24-AUG-2001	
Mus musculus HPH1 (Hfh1) gene, partial cds.	
AF154426	
AF154426	
AF154426.1 GI:8347144	
Mus musculus (house mouse)	
Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
1 (bases 1 to 1513)	
Hong H.-K., Noveroske, J.K., Headon, D.J., Liu, T., Sy, M.S.,	
Justice, M.J. and Chakravarti, A.	
The winged helix/forhead transcription factor Foxq1 regulates	
differentiation of hair in satin mice	
Genesis 29 (4), 163-171 (2001)	
21207067	
11309849	
2 (bases 1 to 1513)	
Hong H.-K., Noveroske, J.K., Justice, M.J. and Chakravarti, A.	
Direct Submission	
Submitted (26-MAY-1999) Genetics, Case Western Reserve University,	
10900 Euclid Avenue, Cleveland, OH 44106, USA	
Location/Qualifiers	
1..1513	
/organism="Mus musculus"	
/mol_type="genomic DNA"	
/strain="129"	
/db_xref="taxon:10090"	
<158..>1360	
/gene="Hfh1"	
<158..>1360	
/gene="Hfh1"	
/product="Hfh1"	
158..1360	
/gene="Hfh1"	
/codon_start=1	
/product="Hfh1"	
/protein_id="AAP74524.1"	
/db_xref="GI:8347145"	
/translation="MKLEVPVRAAHGDKMGSDLEGAGSDVPSPLSAGDDSLGSDG	
DCANSPAGSGAGDLEGGERNSSGGPSAOGPBATDDSRTOAGSAGCAGGVGG	
EGARSKPTTRPKPPYSYIALIAIRDSAGRLTLAEINELMGKFPFPRGSGTGW	
NSVHNLSLNCDFVKVLRDSRFRWGNMNPNSBYTPADGVFRERRLSHRTTV	
SASGLRPEEPAGPGAGTQPPAPARSPARSPARSPARSPARSPARSPARSPARSP	
PFPSRRDGSALGVQFPWGAAPCPFLRAYPALLPAPAGGALLPLCAYGASEPTLLSK	
GTEVQPAAPLLAPLSTAPAPKPRGPETAGAAHLYCPLRLPTALQAAAACGPGPHLS	
YVPEVTLA"	

US-10-087-080-32 (1-402) x AF154426 (1-1513)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
 Db 158 ATGAATTTGGAGTGTTCCTCCACGCGAGCCACCGGGGACAAATCGGCGAGGATCTG 217  
 QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40  
 Db 218 GAGGGGGCGGCGAGCAGCGAGTGCATCTCCACTGTCCGCGGTGGTGACGACTCCTTA 277  
 QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr 59  
 Db 278 GCTCAGACGGGAGTGTGCAGCCACAGCCCGCGCGGCGAGCGCGCGGGGATCTG 337  
 QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIlePro 78  
 Db 338 GAAGGTGGCGGCGGAGGAGGATTCGAGTGGCGGCGGCGAGCCCGAGCGTCCGAG 397  
 QY 79 AlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGly 98  
 Db 398 GCACTGATGACAGAGA-----ACGACGAGCTCCGCGGCGAGGCGCGTGC 442  
 QY 99 AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLys 118  
 Db 443 CGGGGCGGCGTGGCGGCGGAGCGGCGGCGAGCCGCTACACGCGGCGGCCAG 502  
 QY 119 ProProTyrSerTyrIleAlaAlaIleAlaMetAlaIleAlaArgAspSerAlaGlyArg 138  
 Db 503 CCCCATACTCTCATCTGCTCTCATCTGCCATGCGCATCCCGGACTCCGCGGGCGGACGC 562  
 QY 139 LeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySer 158  
 Db 563 CTGACACTGGCGAGATCAAGATCACTCATGGGCAAGTTCCCTTTTCCGGGGGAGC 622  
 QY 159 TyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLys 178  
 Db 623 TACACGGGCTGGCGCAACTCCGTGGCCACACCTCTCGCTCAACGACTGTTTCGTCAG 682  
 QY 179 ValLeuArgAspProSerArgProTyrGlyLysAspHyrTyrTyrMetLeuAsnProAsn 198  
 Db 683 GTGCTGGCGGACCCCTCGCGCCCTGGGGCAAGACAACTACTGGATGCTCAACCCCAAC 742  
 QY 199 SerGluTyrThrPheAlaAspGlyValPheArgArgArgArgLysArgLeuSerHisArg 218  
 Db 743 AGCGAATACACTTCGCGCAGCGGGTCTTCGCGCGCGCGCAAGCGCTCAGCCACCGG 802  
 QY 219 AlaProValProAlaProGlyLeuArgProGluAlaProGlyLeuProAla---Ala 237  
 Db 803 ACCACAGTCTCCGGCTCGGCGCGGAGGAGCCCGACCGGACCTGCGGGGACC 862  
 QY 238 ProProProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGlnGlu 257  
 Db 863 CCGAGCCCGCGCGCGCGCGCTCTCCCGGATCGCGGCTCGCGGCTCGCCAGGAG 922  
 QY 258 GluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArg 277  
 Db 923 GAGCGCTCCAGCCCTCGCGAGCAAGTTCTCCAGCTCTTCGCGATCGACAGCATCTCAGC 982  
 QY 278 LysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTyrGly 297  
 Db 983 AAGCTTTTTCGAGCGCGCGCGGAGCGGCGACTCTGCGCTCTGGGGGTGCGACTACCTTGGGGC 1042  
 QY 298 AlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArg 317  
 Db 1043 CGCGCTCCCTCGCGCGCTCGCGGCTATCCCGCGCTCTTCGCGGCGCGCGCGGTGC 1102  
 QY 318 AlaLeuLeuProLeuCysAlaTyrGlyAlaGlyLysProAlaArgLeuGlyAlaArgGlu 337  
 Db 1103 GCTCTGTACCGCTCTGTGCTTACCGGCGCAAGCGAGCGCTACGCTGCTGGCGTGGCGGGG 1162  
 QY 338 AlaGluValProThrAlaProLeuLeuAlaProLeuLeuAlaAlaAlaPro 357  
 Db 1163 ACCGAGGTGACCGCGCGCG---CCCTTCTGTGCGCGCCCTCTCCACCGCGGCTCCA 1219

358 AlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeuArg 376  
 Db 1220 GCCACGCCATTCGAGTTCGGAGACCGCGCGCGCGCGACCTGACTGCCCCCTACGG 1279  
 QY 377 LeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrPro 396  
 Db 1280 CTGCCACGCGCTCGAGCGGCGAGCGGCTCGGCTCCGCTCCGACCTGTCTCTACCG 1339  
 QY 397 ValGluThrLeuLeuAla 402  
 Db 1340 GTGGAGACTCTGCTAGCT 1357  
 RESULT 6  
 BC047155 2046 bp mRNA linear ROD 07-OCT-2003  
 LOCUS Mus musculus forkhead box Q1, mRNA (CDNA clone MGC:54661  
 DEFINITION IMAGE:6489152), complete cds.  
 BC047155  
 BC047155.1 GI:28502981  
 VERSION MGC.  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 (bases 1 to 2046)  
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
 Stapleton,M., Soares,M.B., Donald,M.F., Casavant,T.L.,  
 Schettler,T.E., Brownstein,M.J., Udwin,T.B., Toshikiyuki,S.,  
 Carrinon,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
 Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,  
 McKernan,K.C., Malek,J.A., Gunaratne,P.H., Richards,S.,  
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
 Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
 Fahay,J., Helton,E., Kettner,M., Madan,A., Rodriguez,S.A.,  
 Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
 Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalieu,D.E.,  
 Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932  
 REFERENCE  
 2 (bases 1 to 2046)  
 Strausberg,R.  
 Direct Submission  
 Submitted (21-FEB-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
 Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,  
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,  
 Madeno,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,  
 Haddou,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,  
 McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,

Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 100 Row: n Column: 18  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6754191.

## FEATURES

## source

Location/Qualifiers  
 1..2046  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="MGC:54661 IMAGE:6489152"  
 /tissue\_type="Mammary tumor, MMTV-LTR/INT3 model, 5 month old mouse, taken by biopsy."  
 /clone\_lib="NCI CGAP Mam2"  
 /lab\_host="DH10B"  
 /notes="Vector: pCMV-SPORT6"

## gene

1..2046  
 /gene="Foxq1"  
 /notes="synonym: FHF-1"  
 /db\_xref="LocusID:15220"  
 /db\_xref="MGI:1298228"  
 481..1683

## CDS

/codon\_start=1  
 /product="forkhead box Q1"  
 /protein\_id="AAH47155.1"  
 /db\_xref="GI:28502982"  
 /db\_xref="LocusID:15220"  
 /translations="MKLEVFVFAAHGDKVGLSGASSDVPFLSAGDDSLGSDG  
 DCAANSFPAAGSGADLGGGGNRSQGPDAVDSDRTQASAPGAGVGGG  
 EGARSPYTRRPKPYSYTALAMARDAGAGLTLAEINLMGKFFPRGSYTGWR  
 NSVHNLINDCFKVLRPSPWKNQDNYMLNPSEYTFADGVPRRRKRLSHRTTV  
 SASGLRPERAPPGAGTPOAPAAASPIARPAPQERSSPASXFSFADSLSK  
 PPSRDGDSALGVLPWGAAPCPPIRAYPALLPAAPGALLPLCAYGASEPTILASR  
 GTVQPAAPLLAPLSTAPAPKPPFPETAGAHLYCPLLELTALQAAACGPGHLS  
 YPVETLLA"  
 817..1089

## misc\_feature

/note="FH; Region: FORKHEAD"  
 /db\_xref="CDD:smart00339"

## ORIGIN

## Alignment Scores:

Pred. No.: 8.39e-36 Length: 2046  
 Score: 1597.00 Matches: 320  
 Percent Similarity: 82.27% Conservative: 14  
 Best Local Similarity: 78.82% Mismatches: 62  
 Query Match: 75.22% Indels: 10  
 DB: 10 Gaps: 6

US-10-087-080-32 (1-402) x BC047155 (1-2046)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
 Db 481 ATGAATTGGAGGTGTTTCGTCACCGCGCAGCCACCGGGGACAAAATGGCAGCGATCTG 540  
 QY 21 GluGlyAlaGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40  
 Db 541 GAGGGGCGCGGACGACGCGATGTCATCTCCATGTCCTGCGCGCTGTCACGACTCCTTA 600  
 QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaAspThr 59  
 Db 601 GGCTCAGCGGGACTGTGCAGCCACACAGCCGCGCGGCGGACGCGCGGGGATCTG 660  
 QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluAlaIlePro 78  
 Db 661 GAAGGTGGCGCGGAGAGAAATTCAGTGTGGCGGCGGAGCCCAAGACGCTCCGAG 720  
 QY 79 AlaAlaAlaAlaAlaAlaValAlaGluAlaGlyAlaAlaGlyAlaGlyProGly 98  
 Db 721 GCAACTGATGACACAGA-----ACGCGAGCCTCCGCGGCGAGGCGGTGC 765

## RESULT 7

AL589738

LOCUS

DEFINITION

Mouse DNA sequence from clone RP23-322J11 on chromosome 13,

complete sequence.

ACCESSION

AL589738

VERSION

AL589738.6

GI:15422196

KEYWORDS

HTG.

187709 bp DNA linear ROD 29-JUN-2002  
 Mouse DNA sequence from clone RP23-322J11 on chromosome 13,  
 complete sequence.  
 AL589738  
 AL589738.6 GI:15422196  
 HTG.

QY 99 AlaGlyGlyAlaGlySerGlyGluGlyAlaAlaSerLysProThrThrArgA-gProLys 118  
 Db 766 CGCGGCGCGGTGGCGGCGGAGGGCGCGCAGCAAGCGTACACGCGCGGCCCAAG 825  
 QY 119 ProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArg 138  
 Db 826 CCCCATCTCTCTACATCGCTCTCATCGCATCGCATCGCATCGCGGCGGAGCGC 885  
 QY 139 LeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySer 158  
 Db 886 CTGACACTGGCGGAGATCAACGAGTACTCTATGGGCAAGTTCCTTTTCCGGGCGCAG 945  
 QY 159 TyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLys 178  
 Db 946 TACACGGGCTGGCGCAACTCGTGGCGCCACAACCTCTCGTCAACGACTGTTTCGTCAG 1005  
 QY 179 ValLeuArgAspProSerArgProTyrGlyLysAsnAsnTyrTrpMetLeuAsnProAsn 198  
 Db 1006 GTGCTGGCGACCCCTCGCGGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1065  
 QY 199 SerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArg 218  
 Db 1066 AGCGAATACACTTCGCGGCGCGGCTCTTCGCGCGCGGCAAGCGCTCAGCACCGG 1125  
 QY 219 AlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla---Ala 237  
 Db 1126 ACCACAGTCTCCGCGTCCGCGTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1185  
 QY 238 ProProProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGluGlu 257  
 Db 1186 CCGCAGCCCGCGCGCGCGCGCTCTCCGCGCTCGCGGCTCGCGGCTCGCGGAG 1245  
 QY 258 GluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArg 277  
 Db 1246 GAGCGCTCCAGCCCTCGCGAGCAAGTCTTCAGCTCTTCGCGCATCGACAGCATCTCAGC 1305  
 QY 278 LysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGly 297  
 Db 1306 AAGCTTTTCGCGCGCGCGCGCGGAGCTCGGCTCGGCGGTGCGAGTACCCCTGGGCG 1365  
 QY 298 AlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArg 317  
 Db 1366 CCGCGTCCCTCGCGCGCGCTCGCGGCTATCCCGCGCTCTTCGCGGCGCGCGGTGGC 1425  
 QY 318 AlaLeuLeuProLeuCysAlaTyrGlyValGluProAlaAlaArgLeuGlyAlaArgGlu 337  
 Db 1426 GCTCTGCTACCGCTCTGTGCTTACGCGGCGCAGCGAGCTACGCTGCTGGGTGCGCGGG 1485  
 QY 338 AlaGluValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaPro 357  
 Db 1486 ACCGAGGTGCGCGCGCGCGCG---CCGCTTCTGCTGCGCGCGCTCTCCACGCGGCTCCA 1542  
 QY 358 AlalysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeuArg 376  
 Db 1543 GCCAAGCCATTCGAGGTCGCGGAGACCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGG 1602  
 QY 377 LeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrPro 396  
 Db 1603 CTGCCACCGCGCTTCAGCGCGCAGCGGCTCGCGTCCGCGTCCGCGTCCGCGTCTCTTACCCG 1662  
 QY 397 ValGluThrLeuLeuAla 402  
 Db 1663 GTGGAGACTCTGCTAGCT 1680



DEFINITION Mus musculus fork head transcription factor (Hfh-1L) gene, complete.

ACCESSION AF010405

VERSION AF010405.2 GI:6716869

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4763)

AUTHORS Frank, S. and Zoll, B.

TITLE Mouse HNF-3/fork head homolog-1-like gene: structure, chromosomal location, and expression in adult and embryonic kidney

JOURNAL DNA Cell Biol. 17 (8), 679-688 (1998)

MEDLINE 98392851

PUBMED 9726250

REFERENCE 2 (bases 1 to 4763)

AUTHORS Frank, S. and Zoll, B.

TITLE Direct Submission

JOURNAL Submitted (19-JUN-1997) University of Goettingen, Institute of Human Genetics, Gosselerstr 12d, Goettingen 37073, Germany

REFERENCE 3 (bases 1 to 4763)

AUTHORS Pasche, B., Bieller, A. and Zoll, B.

TITLE Direct Submission

JOURNAL Submitted (19-JAN-2000) University of Goettingen, Institute of Human Genetics, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany

REMARK Sequence update by submitter

COMMENT On Jan 20, 2000 this sequence version replaced gi:2997586.

FEATURES

source

1..4763

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="129"

/db\_xref="taxon:10090"

/chromosome="13"

/tissue\_type="kidney"

<2467..33669

/gene="Hfh-1L"

/note="HNF-3/fork head homologue-1 like"

<2467..33669

/gene="Hfh-1L"

/product="fork head transcription factor"

2467..3669

/gene="Hfh-1L"

/note="member of HNF-3/fork head transcription factor family"

/codon\_start=1

/product="fork head transcription factor"

/protein\_id="AAC12973.2"

/db\_xref="GI:6716870"

/translation="MKLEVFVPRAAHGDWGSDEGAGSDVPSPLSAGDDSLGSDG  
DCANSPAGSGAGDLEGGGGRNNSGGSSAQDGPATDDSRDQASAGPCAGGVGG  
EGARSKFYTRPRKPPYSYALIAIARDSGRLTLAEINEYLMGFPPFRSGYTGWR  
NSVHNLSLNDCKVKLRDPSRWKDNWMLNPSEYTFADGVFRRKRSLHRTV  
SAGLRDEEPAPGPTQAPAPSPARSPIARSPARBERSPSKFSFSAIDSLSK  
PFSRRDGDALGVQFWGAPCPPLRAYPALPAPGALLPLCAYGASEPTLLASR  
GTEVQPAAPLLAPLSTAAKPFPRGPETAGAAHLYCPLRLPTALQAAAACGPGPHLS  
YFVETLLA"

2782..3123

/gene="Hfh-1L"

/note="encodes DNA binding domain"

ORIGIN

Alignment Scores:

Prod. No.: 2,66e-35 Length: 4763

Score: 1589.00 Matches: 319

Percent Similarity: 82.02% Conservative: 14

Best Local Similarity: 78.57% Mismatches: 63

Query Match: 74.85% Indels: 10

DB: 10 Gaps: 6

US-10-087-080-32 (1-402) x AF010405 (1-4763)

QY 1 MetLysIeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20

Db 2467 ATGAATTTGAGGGTTCGTCACGCGCAGCCACGGGACAAATAATGGCAGCGCATCTG 2526

QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40

Db 2527 GAGGGGCGCGCAGCAGCGAGCGTCCATCTCCATGTCCGCGGTGTGTGACGACTCCTTA 2586

QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr 59

Db 2587 GGCTCAGCAGGGGACTGTGAGCCCAACAGCCCGCGCGGCGGACGCGCGGGGATCTG 2646

QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluAlaIlePro 78

Db 2647 GAAGTGTGGCGCGCAGAGAAATTCAGTGGCGGCGGAGCGCCCAAGACGGTCCGAG 2706

QY 79 AlaAlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGly 98

Db 2707 GCAACTGATGACAGCAGA-----ACGCGAGCCCTCCGCGGAGCGGCGCGTGC 2751

QY 99 AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgProLys 118

Db 2752 CGCGCGCGGTGGCGCGCGGAGGCGCGCAGCAAGCCGTACACGCGCGGCGCCCAAG 2811

QY 119 ProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyArg 138

Db 2812 CCCCACATCTCTATCGCATCGCTCTCATCGCATCGGACTCCGCGGCGGAGCGC 2871

QY 139 LeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySer 158

Db 2872 CTGACACTGCGCGAGATCAACGAGTACTCATGGCAAGTTCCTCTTTCCGCGGCGC 2931

QY 159 TyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLys 178

Db 2932 TACACGGGCTGGCGCAATCTCGTGGCGCAACACCTCTCGTCAACGACTGTTTCGTCAG 2991

QY 179 ValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsn 198

Db 2992 GTGCTGGCGRACCCCTCGGCGCTCGGGCAAGCAACTACTGATGCTCAACCCCAAC 3051

QY 199 SerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArg 218

Db 3052 AGCGAATACACTTTCGCGCAGCGGGTCTTCGCGCGCGCGCAAGCGCTCAGCACCGG 3111

QY 219 AlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla---Ala 237

Db 3112 ACCACAGTCTCCGCGTCCGGCTCGCGCGGAGGAGACCCACCGGACTGCGCGGACC 3171

QY 238 ProProAlaProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGlnGlu 257

Db 3172 CCGCAGGCGCGCGCGCGCGCTCTCCCGATCGCGCGCTCGCGGCTCGCCAGAG 3231

QY 258 GluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArg 277

Db 3232 GAGCGCTCCAGCCCTCGGAGCAATTCACAGTCTTCGCGCATCGACAGCATCTCTCAGC 3291

QY 278 LysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTyrGly 297

Db 3292 AAGCTTTTCGAGCGCGCGCGCGGACTCGGCTCTGGGGGTGACGTACCTCGGGGC 3351

QY 298 AlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArg 317

Db 3352 GCCGCTCCCTCGCGCGCTATCCCGGCTCTTCGCGGCGCGCGCGCGGCTGCGC 3411

QY 318 AlaLeuLeuProLeuCysAlaTyrGlyValaGlyGluProAlaArgLeuGlyAlaArgGlu 337

Db 3412 GCTCTGCTACCGCTCTGTGCTTACGGCGCAGCGAGCTACGCTGCTGGCGTCCGCGGG 3471

QY 338 AlaGluValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaPro 357

Db 3472 ACCGAGGTGCAGCGCGCGCG---CCCTTCTGTGGCGCGCCCTCTCCACCGCGGCTCCA 3528

QY 358 AlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeuArg 376



```

Db      1151  GCGCCAGCAAGCATTTCAGTCCGAGACCGCGCGCGCGCACCTGCTACTGCCCC 1210
Oy      375  LeuAqgLeuProAlaLeuGlnAlaLeuValArgArgProGlyProHisLeuSer 394
Db      1211  CTACGGCTGCCACAGCCCTCGAGGCGCGCGCGCTCGGTCGCGGTCGCGACCTGTCC 1270
Oy      395  TyrProValGluThrLeuLeuAla 402
Db      1271  TACCGGGTGGAGACGCTGTAGCT 1294

RESULT 10
AC119497
LOCUS   Rattus norvegicus clone CH230-521F8, *** SEQUENCING IN PROGRESS
DEFINITION
***, 4 unordered pieces.
AC119497
VERSION 15184 bp DNA linear HTG 20-NOV-2002
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE  Rattus norvegicus (Norway rat)
ORGANISM
Mammalia; Eutheria; Rodentia; Chordata; Vertebrata; Euteleostomi;
Rattus.
REFERENCE
1 (bases 1 to 195184)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgopoulos,E., Geier,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Haviak,P., Haves,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebowitz,Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegged,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Packerleh,O., Okwuonu,G., Olarpunsgoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plapper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojass,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smays,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

```

```

Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 195184)
Worley,K.C.
Direct Submission
Submitted (27-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 195184)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23195197.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWQI
Center clone name: CH230-521F8
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 158913 bases at least Q40
Consensus quality: 162109 bases at least Q30
Consensus quality: 164133 bases at least Q20
Estimated insert size: 161675; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 16506: contig of 16506 bp in length
* 16507 16606: gap of unknown length
* 16607 180638: contig of 164032 bp in length
* 180639 180738: gap of unknown length
* 180739 182447: contig of 1709 bp in length
* 182448 182547: gap of unknown length
* 182548 195184: contig of 12637 bp in length.
* Location/Qualifiers
* 1..195184
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clones="CH230-521F8"
* 466..1181
* /note="clone boundary
* clone end:T7
* site:
* end sequence:BZ266854"
* 1768..2639
misc_feature
source
FEATURES
misc_feature
end sequence:BZ266854"
1768..2639

```

```

/note="clone boundary
clone_end:Sp6
site:
end_sequence:BZ267695"
misc_feature
16607..17640
/note="wgs_end_extension
clone_end:Sp6"
misc_feature
71528..72576
/note="wgs_end_extension
clone_end:Sp6"
misc_feature
76693..78955
/note="wgs_end_extension
clone_end:Sp6"
misc_feature
86077..87329
/note="wgs_end_extension
clone_end:Sp6"
misc_feature
106805..108354
/note="wgs_end_extension
clone_end:Sp6"
misc_feature
129622..130958
/note="wgs_end_extension
clone_end:Sp6"

ORIGIN
Alignment Scores:
Pred. No.: 1.45e-32 Length: 195184
Score: 1534.00 Matches: 311
Percent Similarity: 79.90% Conservative: 15
Best Local Similarity: 76.23% Mismatches: 68
Query Match: 72.28% Indels: 14
DB: 2 Gaps: 7

US-10-087-080-32 (1-402) x AC119497 (1-195184)
QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db 136153 ATGAATTTGGAGTATTTGCCCCACGCGCCGACGCGGACAGATGGCGAGTGCCTG 136212
QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40
Db 136213 GAGGGGGCGGCGAGCAGCGAGTGCATCTCGCTGTCTCGGGGTGGGAGCACTCTCTTA 136272
QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr 59
Db 136273 GGCTCTGACGGGAGCTGTGTACACCAACAGCCCGCGCGGCGAGCGCCGTGATCTG 136332
QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIlePro 78
Db 136333 GAAGCGGCGGCGGCGAGAGGATTCGAGTGGCGGGCGGAGCAACCCAGAC----- 136383
QY 79 AlaAlaAlaAlaAlaValValAlaGluGlyAla-----GluAlaGlyAlaAlaGly 96
Db 136384 -----GATCCCGAGGTGACCGATGGCAGCAGCAGACGCGCCTCCCGGTGGGG 136431
QY 97 ProGlyAlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArg 116
Db 136432 CCGTGGCGGGGAGCGGTGGCGGGGTGAGCGCGGCGGCGGCGGAGCAGCCGTACAGCGGGG 136491
QY 117 ProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleAlaArgAspSerAlaGly 136
Db 136492 CCCAAGCCCCCTACTCTTACATCGCATCTCATCGCATCGCCATCCGCGACTCCCGCGGGC 136551
QY 137 GlyArgLeuThrIleuAlaGluIleAlaGluTyrLeuMetGlyLysPheProPhePheArg 156
Db 136552 GGACGGCTGACGTGGCGGAGATCAACGAGTACCTCATCGGGCAAGTTCCTCTTTTCGGG 136611
QY 157 GlySerTyrThrGlyTrpArgAsnSerValArgHisAsnLeuSerIleuAsnAspCysPhe 176
Db 136612 GGCAGCTACACGGCTGGCGCAACTCCGTGCGCCACACACTCTCGCTCAACGACTGTTC 136571
QY 177 ValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsn 196
Db 136572 GTCAAGGTGCTGGCGGACCCCTCGCGGCGCTGGGGCAAGGACAAATTACTGGATGCTCAAC 136731

QY 197 ProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgArgArgArgLeuSer 216
Db 136732 CCACACGCGCAATACACCTTCGCCCGCGGGGTCTTCGCCCGCGCGCGCAAGCGCTCAGC 136791
QY 217 HisArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla 236
Db 136792 CACCGGACCACAGTCTCCGCTACGCGGCTACGCGCGGAGGAGGAGCCACCCGAGACTGG 136851
QY 237 ---AlaProProAlaProAlaProAlaProAlaProArgMetArgSerProAlaArg 255
Db 136852 GGGACCCCGCAGCCCGCGCCACCCCGCGGTCTCTCCCAATCGCGGCTCGCCGCTCGC 136911
QY 256 GlnGluGluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIle 275
Db 136912 CAGGAGGAGGCTCCAGCCCGCGGAGCAAGTCTCCAGCTCTTCGCCATCGACAGCATC 136971
QY 276 LeuArgLysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGln 295
Db 136972 CTCAGCAAGCCGTTTCGACGCGCGCGGACCGCGGCTCTGGGGGTGAGCTACCC 137031
QY 296 TrpGlyAlaAlaProCysProLeuProAlaPheProAlaLeuLeuProAlaAlaPro 315
Db 137032 TGGAGCGCTGCTCCCTGCGCGCGCTGCGCGCTATCCCGGCTCTCTCCGCGCTGCTCC 137091
QY 316 CysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAla 335
Db 137092 GCGCGTGCCTGCTGCGCTCTGTGTACGCGCGCGGCGGAGCCACGCTGTGGCGTGG 137151
QY 336 ArgGluAlaGluValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAla 355
Db 137152 CGCGGGCGCGAGGTGAGCGCGCGGG---CCCTGTGCTGGCGCGCTCTCCACCGGG 137208
QY 356 AlaProAlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysPro 374
Db 137209 GCCCCAGCAAGCAATTCGAGTTCGCGAGACCGCGCGCGCGGCGGCACTGTACTGCC 137268
QY 375 LeuArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSer 394
Db 137269 CTACGGTGGCCACGCGCTTGCAGGGCGCGCGGCTCGGCTCGGAGTCCGAGTCCGACCTGTCC 137328
QY 395 TyrProValGluThrLeuLeuAla 402
Db 137329 TACCGGTGGAGACGCTGTAGCT 137352

RESULT 11
RATHFH1
LOCUS
DEFINITION
Rattus norvegicus HNF-3/forkhead homolog-1 (HNF-1) mRNA, complete cds.
ACCESSION
L13201
VERSION
L13201.1 GI:951067
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1760)
Clevidence,D.E., Overdier,D.G., Tao,W., Qian,X., Pani,L., Lai,E.
and Costa,R.H.
Identification of nine tissue-specific transcription factors of the
hepatocyte nuclear factor 3/forkhead DNA-binding-domain family
Proc. Natl. Acad. Sci. U.S.A. 90 (9), 3948-3952 (1993)
93248207
PUBMED
7683413
COMMENT
On Aug 22, 1995 this sequence version replaced gi:550512.
Original source text: rat.
Location/Qualifiers
1..1760
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
FEATURES
source

```

US-10-087-080-32 (1-402) x RATHFH1 (1-1760)

QY	1	MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu	20
Db	207	ATGAATTTGGAGTG-ATTGCTCCACCGCAGC-CACGGGGACAAATGGCAGCATCTG	364
QY	21	GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu	40
Db	265	GAGGGGCGGCAGCAGCAGTGCATCTCCACTGTCGCGGGTGGTAGACTCTCTTA	324
QY	41	GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr	59
Db	325	GGCTCAGACGGGACTGTGCAGCAACACCGCGGGCGGCAGCGCGCGGGGATCTG	384
QY	60	Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIlePro	78
Db	385	GAAGGTGGCGCGGAGAGGATTCGATGTGGCGGGCGAGCACCCCAAGAC-----	435
QY	79	AlaAlaAlaAlaAlaValValAlaGluGlyAla-----GluAlaGlyAlaAlaGly	96
Db	436	-----GATCCCGAGGTGACCGGTATGCAGCAGAACGCGAGCGCTCCCGGTGGG	483
QY	97	ProGlyAlaGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArg	116
Db	484	CCGTGCGCGGAGCGGTGGCGCGGTGAGGCGCGCGCACAGCGGTACACGCGCGG	543
QY	117	ProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGly	136
Db	544	CCCAAGCCCCCTACTCTCTACATCGCACATCGCCATCGGCATCCCGCGACTCCGCGGGC	603
QY	137	GlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArg	156
Db	604	GGAGCGCTGACGCTGGCCGAGATCAACGAGTAGTACCTCATGGGCAAGTTCCTCTTTCCGG	663
QY	157	GlySerTyrThrGlyTyrPargAsnSerValArgHisAsnLeuSerLeuAsnAspCysPhe	176
Db	664	GGCAGCTACACGGGCTGGCGCACTCCGTGCGCCACAACCTCTCGCTCAACGACTGTTTC	723
QY	177	ValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsn	196
Db	724	GTCAAGGTGTGGCGCACCCCTCGGGCCCTGGGCAAGGCAATTAATCGATGCTCAAC	783

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Melarini, J., Merens, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

# TITLE

JOURNAL  
 Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE AUTHORS

3 (bases 1 to 182152)  
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,  
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,  
 Campopiano, A., Castle, A., Choapel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domano, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kahn, L., Karatas, A.,  
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Melarini, J., Merens, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

## TITLE JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 4, 2000 this sequence version replaced gi:7655991.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIPR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L10103

Center clone name: 403\_I\_20

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 173395 bases at least Q40

Consensus quality: 177417 bases at least Q30

Consensus quality: 179105 bases at least Q20

Insert size: 179000; agarose-1p

Insert size: 180452; sum-of-contigs

Quality coverage: 4.8 in Q20 bases; agarose-1p

Quality coverage: 4.7 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 18 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2222: contig of 2222 bp in length

\* 2223 2222: gap of 100 bp

\* 2323 3785: contig of 1463 bp in length

\*

\* 3786 3885: gap of 100 bp  
 \* 3886 6733: contig of 2868 bp in length  
 \* 6734 6754: gap of 100 bp  
 \* 6754 10036: contig of 3183 bp in length  
 \* 10037 10136: gap of 100 bp  
 \* 10137 14537: contig of 4401 bp in length  
 \* 14538 14637: gap of 100 bp  
 \* 14638 21177: contig of 6540 bp in length  
 \* 21178 21277: gap of 100 bp  
 \* 21278 30240: contig of 8963 bp in length  
 \* 30241 30340: gap of 100 bp  
 \* 30341 40123: contig of 9783 bp in length  
 \* 40124 40223: gap of 100 bp  
 \* 40224 47939: contig of 7776 bp in length  
 \* 48000 48099: gap of 100 bp  
 \* 48100 58235: contig of 10136 bp in length  
 \* 58236 58335: gap of 100 bp  
 \* 58336 67041: contig of 8706 bp in length  
 \* 67042 67141: gap of 100 bp  
 \* 67142 75921: contig of 8780 bp in length  
 \* 75922 76021: gap of 100 bp  
 \* 76022 86528: contig of 10506 bp in length  
 \* 86529 86627: gap of 100 bp  
 \* 86628 101334: contig of 14707 bp in length  
 \* 101335 101434: gap of 100 bp  
 \* 101435 117630: contig of 16196 bp in length  
 \* 117631 117730: gap of 100 bp  
 \* 117731 138293: contig of 20563 bp in length  
 \* 138294 138393: gap of 100 bp  
 \* 138394 157841: contig of 19448 bp in length  
 \* 157842 157941: gap of 100 bp  
 \* 157942 182152: contig of 24211 bp in length.

## FEATURES Location/Qualifiers

source  
 1..182152  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6"  
 /clone="RP11-403I20"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 misc\_feature  
 1..2222  
 /notes="assembly\_fragment"  
 misc\_feature  
 2323..3785  
 /notes="assembly\_fragment"  
 misc\_feature  
 3886..6753  
 /notes="assembly\_fragment"  
 misc\_feature  
 6854..10036  
 /notes="assembly\_fragment"  
 misc\_feature  
 10137..14537  
 /notes="assembly\_fragment"  
 misc\_feature  
 14638..21177  
 /notes="assembly\_fragment"  
 misc\_feature  
 21278..30240  
 /notes="assembly\_fragment"  
 misc\_feature  
 30341..40123  
 /notes="assembly\_fragment"  
 misc\_feature  
 40224..47939  
 /notes="assembly\_fragment"  
 misc\_feature  
 clone\_end:Sp6  
 vector\_side:left  
 48100..58235  
 /notes="assembly\_fragment"  
 misc\_feature  
 58336..67041  
 /notes="assembly\_fragment"  
 misc\_feature  
 67142..75921  
 /notes="assembly\_fragment"  
 misc\_feature  
 76022..86527  
 /notes="assembly\_fragment"  
 misc\_feature  
 86628..101334  
 /notes="assembly\_fragment"  
 misc\_feature  
 101435..117630  
 /notes="assembly\_fragment"

```

misc_feature 11731..138293
/ note="assembly_fragment"
138394..157841
/ note="assembly_fragment"
157942..182152
/ note="assembly_fragment"
clone_end.T7
vector_side:right"

ORIGIN
Alignment Scores:
Pred. No.: 3.39e-28 Length: 182152
Score: 1372.00 Matches: 255
Percent Similarity: 99.22% Conservatives: 2
Best Local Similarity: 99.22% Mismatches: 0
Query Match: 64.63% Indels: 0
DB: 2 Gaps: 0

US-10-087-080-32 (1-402) x AC067929 (1-182152)
QY 146 GluTyrLeuMetGlyLysPheProPheArgGlySerTyrThrGlyTPArgAsnSer 165
DB 10036 GAGTACTCATGGCGAAGTTCCTTTTCGGGGAGCTACACGGCTGGCGCACTCC 9977
QY 166 ValArgHisLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArg 185
DB 9976 GTGGCCACACCTTCGCTCACGACTGCTTCAGGTGTCTGCGGACCCCTCGCG 9917
QY 186 ProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAsp 205
DB 9916 CCTTGGGGAGGACACTACTGATGCTCAACCCCAAGCAGATACACTTCGGCGAC 9857
QY 206 GlyValPheArgArgAArgLysArgLeuSerHisArgAlaProValProAlaProGly 225
DB 9856 GGGGTCTTCGCCCGCGCGAGCGCTCAGCCACCGCGCGCGTCCCGCGCGCGG 9797
QY 226 LeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaPro 245
DB 9796 CTGGGGCCGAGGAGGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 9737
QY 246 AlaSerProArgMetArgSerProAlaArgGlnGluArgAlaSerProAlaGlyLys 265
DB 9736 GCCTCGCGCGCATGCGTCCCGCGCGCGCGCGCGAGGAGCGCGCAGCCCGCGG 9677
QY 266 PheSerSerPheAlaAlaAspSerIleuArgLysProPheArgSerArgArgLeu 285
DB 9676 TTCTCCAGCTCCTTCGCGCATCGACGATCTCGCGCAAGCCCTTCGCGACCGCGC 9617
QY 286 ArgAspThrAlaProGlyThrThrLeuGlnTyrGlyAlaAlaProCysProLeuPro 305
DB 9616 AGGACACGCGCCCGCGGACGACGCTTCAGTGGGCGCGCGCGCGCTGCGCGTCC 9557
QY 306 AlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyr 325
DB 9556 CGCTTCCCGCGCTCCTCCCGCGCGCGCTTCGAGGCGCGCTGCTGCGCTCGCGTAC 9497
QY 326 GlyAlaGlyGluProAlaAlaArgLeuGlyValArgGluAlaGluValProThrAlaPro 345
DB 9496 GCGCGGCGGAGCGCGCGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9437
QY 346 ProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysProLeuArgGlyProAla 365
DB 9436 CCGCTCTCTGCTTGACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9377
QY 366 AlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaLeu 385
DB 9376 CCGCGGCGCGCGCGCGCGCTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9317
QY 386 ValArgArgProGlyProHisLeuSerTyrProValGluThrLeuLeuAla 402
DB 9316 GTCCGCGCGCGCGCGCGCGCTGCGGTACCGCGTGGAGCGCTCTCTAGCC 9266

```

RESULT 13

AC012350/c  
LOCUS  
DEFINITION

AC012350 178168 bp DNA linear HTG 01-APR-2000  
Homo sapiens clone RP11-16N9, WORKING DRAFT SEQUENCE, 16 unordered  
pieces.

AC012350  
VERSION  
HTG: HTGS\_PHASE1, HTGS\_DRAFT.  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
Homo sapiens  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 178168)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens, clone RP11-16N9  
Unpublished  
2 (bases 1 to 178168)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boguski,M., Collins,S., Collymore,A.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehotzky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McWan,P., McMurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Submitted (25-OCT-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 1, 2000 this sequence version replaced g1:6479001.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBK

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3593

Center Clone name: 16.N.9

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 172070 bases at least Q40

Consensus quality: 174479 bases at least Q30

Consensus quality: 175597 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 178668; sum-of-contigs

Quality coverage: 6.4 in Q20 bases; agarose-fp

Quality coverage: 6.2 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 16 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1694: contig of 1694 bp in length  
\* 1695 1794: gap of 100 bp  
\* 1795 4437: contig of 2643 bp in length  
\* 4438 4537: gap of 100 bp  
\* 4538 6644: contig of 2107 bp in length  
\* 6645 6744: gap of 100 bp  
\* 6745 10092: contig of 3348 bp in length  
\* 10093 10192: gap of 100 bp



McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,  
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:7144942.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WtBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L3560  
Center clone name: 16\_J\_9  
-----  
\* NOTE: This record contains 73 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
1  
944: contig of 944 bp in length  
1  
945  
1044: gap of 100 bp  
1045  
1991: contig of 947 bp in length  
1992  
2091: gap of 100 bp  
2092  
3007: contig of 916 bp in length  
3008  
3107: gap of 100 bp  
3108  
4037: contig of 930 bp in length  
4038  
4137: gap of 100 bp  
4138  
5056: contig of 919 bp in length  
5057  
5156: gap of 100 bp  
5157  
6079: contig of 923 bp in length  
6080  
7103: contig of 924 bp in length  
7104  
7203: gap of 100 bp  
7204  
8136: contig of 933 bp in length  
8137  
8236: gap of 100 bp  
8237  
9166: contig of 930 bp in length  
9167  
9266: gap of 100 bp  
9267  
10167: contig of 901 bp in length  
10168  
10267: gap of 100 bp  
10268  
11212: contig of 945 bp in length  
11213  
11312: gap of 100 bp  
11313  
12250: contig of 938 bp in length  
12251  
12350: gap of 100 bp  
12351  
13277: contig of 927 bp in length  
13278  
13377: gap of 100 bp  
13378  
14275: contig of 898 bp in length  
14276  
14375: gap of 100 bp  
14376  
15281: contig of 906 bp in length  
15282  
15381: gap of 100 bp  
15382  
16299: contig of 918 bp in length  
16300  
16399: gap of 100 bp  
16400  
17341: contig of 942 bp in length  
17342  
17441: gap of 100 bp  
17442  
18359: contig of 918 bp in length  
18360  
18459: gap of 100 bp  
18460  
19396: contig of 937 bp in length  
19397  
19497: gap of 100 bp  
19497  
20457: contig of 961 bp in length  
20458  
20557: gap of 100 bp

20558  
21482: contig of 925 bp in length  
21483  
21582: gap of 100 bp  
21583  
22515: contig of 933 bp in length  
22516  
22615: gap of 100 bp  
22616  
23514: contig of 899 bp in length  
23515  
23614: gap of 100 bp  
23615  
24530: contig of 916 bp in length  
24531  
24631: gap of 100 bp  
24631  
25553: contig of 923 bp in length  
25554  
25653: gap of 100 bp  
25654  
26595: contig of 942 bp in length  
26596  
26695: gap of 100 bp  
26696  
27272: contig of 932 bp in length  
27272  
27628: gap of 100 bp  
27628  
28660: contig of 933 bp in length  
28661  
28760: gap of 100 bp  
28761  
29692: contig of 932 bp in length  
29693  
29792: gap of 100 bp  
29793  
30708: contig of 916 bp in length  
30709  
30808: gap of 100 bp  
30809  
31728: contig of 920 bp in length  
31729  
31828: gap of 100 bp  
31829  
32740: contig of 912 bp in length  
32741  
32840: gap of 100 bp  
32841  
33761: contig of 921 bp in length  
33762  
33804: gap of 100 bp  
33862  
34804: contig of 943 bp in length  
34805  
34804: gap of 100 bp  
34805  
35835: contig of 931 bp in length  
35836  
35935: gap of 100 bp  
35936  
36882: contig of 947 bp in length  
36883  
36982: gap of 100 bp  
36983  
37911: contig of 929 bp in length  
37912  
38011: gap of 100 bp  
38012  
38336: contig of 925 bp in length  
38337  
39036: gap of 100 bp  
39037  
39990: contig of 954 bp in length  
39991  
40090: gap of 100 bp  
40091  
41004: contig of 914 bp in length  
41005  
41104: gap of 100 bp  
41105  
42018: contig of 914 bp in length  
42019  
42118: gap of 100 bp  
42119  
43015: contig of 897 bp in length  
43016  
43115: gap of 100 bp  
43116  
44052: contig of 937 bp in length  
44053  
44152: gap of 100 bp  
44153  
45059: contig of 907 bp in length  
45060  
45159: gap of 100 bp  
45160  
46092: contig of 933 bp in length  
46093  
46192: gap of 100 bp  
46193  
47110: contig of 918 bp in length  
47111  
47210: gap of 100 bp  
47211  
48138: contig of 928 bp in length  
48139  
48238: gap of 100 bp  
48239  
49157: contig of 919 bp in length  
49158  
49257: gap of 100 bp  
49258  
50190: contig of 933 bp in length  
50191  
50290: gap of 100 bp  
50291  
51219: contig of 929 bp in length  
51220  
51319: gap of 100 bp  
51320  
52248: contig of 929 bp in length  
52249  
52348: gap of 100 bp  
52349  
53282: contig of 934 bp in length  
53283  
53382: gap of 100 bp  
53383  
54325: contig of 943 bp in length  
54326  
54425: gap of 100 bp  
54426  
55326: contig of 901 bp in length  
55327  
55426: gap of 100 bp  
55427  
56337: contig of 911 bp in length  
56338  
56437: gap of 100 bp  
56438  
57356: contig of 919 bp in length  
57357  
57457: gap of 100 bp  
57457  
58416: contig of 960 bp in length



DEFINITION Zebrafish DNA sequence from clone DKEY-100F21 in linkage group 20, complete sequence.

ACCESSION BX510653

VERSION BX510653.8 GI:37051088

KEYWORDS HTG.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 62848)

Beasley,H.

Direct Submission

Submitted (29-SEP-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk

zfish-help@sanger.ac.uk

On Sep 29, 2003 this sequence version replaced gi:35761991.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Centre code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)

Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhihong Bao and Sean Eddy, submitted), and those beginning 'arr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml)

DKEY-100F21 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5.

## FEATURES

source

1..62848

Location/Qualifiers

organism="Danio rerio"

mol\_type="genomic DNA"

db\_xref="taxon:7955"

clones="DKEY-100F21"

clone\_lib="DanioKey"

## ORIGIN

Alignment Scores:

Pred. No.: 6,09e-10 Length: 62848

Score: 685.50 Matches: 169

Percent Similarity: 95.03% Conservative: 28

Best Local Similarity: 47.21% Mismatches: 74

Query Match: 32.29% Indels: 87

DB: 5 Gaps: 13

US-10-087-080-32 (1-402) x BX510653 (1-62848)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20

Db 55678 ATGAGTTGGAGTTTC-----TCTGGCAGTCGCTTGTGGACAAGCCGCTGGACCTG 55625

QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40

Db 55624 TGCAGTCACATGCGTGTGAACGTCCTCACCCTGTCCACG-----GAGGAGGAATTG 55571

QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60

Db 55570 GCCTCGGATGGAGACTGCTCTAGCCAAC----- 55544

QY 61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIleProAlaAla 80

Db 55543 -----AGTCCAGAGCCCGCTCCCT----- 55523

QY 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyProGlyAlaGly 100

Db 55522 -----GTTCCGACGGG----- 55511

QY 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120

Db 55510 -----AAGCGAAACCATATACGCTAGACCCCAACCAACCT 55475

QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140

Db 55474 TACTCTTACATGTGCATTTATTCGAATGCCATACCGATTCCACACCGCTCGTCTTACA 55415

QY 141 LeuAlaGluIleAangluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrThr 160

Db 55414 CTTCGTGAATAAAGCAATACTTAATGAAAAATTCCTGTTTTCGGGCGAGCTACACG 55355

QY 161 GlyTrpArgAnSerValArgHisAsnLeuSerLeuAnaspCysPheValLysValLeu 180

Db 55354 GGGTGGCGAAACTCTGACGCGCATAACTGTGCTAAACGACTCTTTTAAAGATTGTTG 55295

QY 181 ArgAspProSerArgProTyrPheLysAspAsnTyrTrpMetLeuAnProAnSerGlu 200

Db 55294 CGAGATCCCTCGCTCCGTGGGCGAAAGACAATTTGGATTGTGAACCCACACAGGAG 55235

QY 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAlaPro 220

Db 55234 TACACTTTTGCAGACGGCGTTTTCGCGGAGGAGGAGCGATCAGT----- 55187

QY 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProPro 240

Db 55186 -----AAGAGATCTCGGCGAGCGGGAATCCCGAGAGCGC 55151

QY 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAla 260

Db 55150 GTGCCCGCGGATGACAGC-----AGGCTTCAGCGCGCAGCAAGC----- 55109

QY 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280

Db 55108 -----GTCTCTTAAGTTTTCAGCTCTTCGCGATAGACATCTTCAGCAACCGTTC 55055

QY 281 -----ArgSerArgArgLeuArgAspThrAlaProGlyThrThr-----LeuGlnTrpGly 297

Db 55054 ATCGGAGCGGAGCAGACAGACAGATGACACTTGTGTTTGTATCCACACAGCTAATGCTGCA 54995

QY 298 AlaAlaPro-----CysProProLeuProAlaPheProAla 309

Db 54994 GCTGCTCTCTCACTTCTGCCAGTTGCCATGGGCTATCGGCCACAGACGAGATTCAGGTC 54935

QY 310 LeuLeuProAlaAlaProCysArgAlaAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGlu 329

Db 54934 -----AGTTCAGGTTGCGAATGCTCTTCCAAATCTACAGGTGTATAACCGCGAC 54884

QY 330 ProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAlaProProLeu 347

Db 54883 ATTTCTAACATC---AGTCGAATGGCGACATACAGAGTTCCCTCCGGGTTG 54833

Search completed: March 31, 2004, 09:01:03

Job time : 4503 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 31, 2004, 04:42:59 ; Search time 444 Seconds

(without alignments)

3846.342 Million cell updates/sec

Title: US-10-087-080-32

Perfect score: 2123

Sequence: 1 MKLEVFVPRAMHGKQSGDL.....AALVRRPGPHLSYPVETLLA 402

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xl  
-O/cn2\_1/USPTO.spcol\_p/US10087080/runat\_29032004.114945.24363/app\_query.fasta\_1.583  
-DB=N\_Geneseq\_25Jan04 -QFNT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=BLOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087080 @CGN 1.1.885 @runat\_29032004.114945.24363 -NCFU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2123	100.0	1209	6 ABQ81557	Abq81557 Gene up-r
2	1412.5	66.5	1760	7 ABT42112	Abt42112 Toxicity
3	790	37.2	658	6 ABQ39023	Abq39023 Oligonucle
C 4	790	37.2	658	6 ABQ39022	Abq39022 Oligonucle
5	488.5	23.0	2487	9 ABT75311	Abt75311 Prostate
6	488.5	23.0	5181	6 ABL68935	Abi68935 Kidney ca
7	487.5	22.0	2271	6 ABZ35021	Abz35021 Human gen
C 8	483	22.8	658	6 ABQ39025	Abq39025 Oligonucle

9	483	22.8	658	6 ABQ39024	Abq39024 Oligonucle
10	472.5	22.3	3946	2 AAZ31671	Aaz31671 Human FKX
11	472.5	22.3	3946	2 AAZ38079	Aaz38079 Human for
12	466.5	22.0	1976	2 AAZ28095	Aaz28095 Human FRE
13	466.5	22.0	2187	7 ABZ34783	Abz34783 Coding se
14	459	21.6	3482	7 ABX76288	Abx76288 Lung canc
15	446.5	21.0	1965	9 ADB31305	Adb31305 Testoster
16	445	21.0	1355	9 ABZ34708	Abz34708 Coding se
17	442.5	20.8	1380	9 ADD29558	Add29558 Human tum
18	429	20.2	1662	2 AAZ31672	Aaz31672 Human FKX
19	429	20.2	1662	2 AAZ38080	Aaz38080 Human for
20	419.5	19.8	2745	7 ABZ20402	Abz20402 Oncofecta
21	419.5	19.8	2753	6 ABT03399	Abt03399 Ovary cel
22	416	19.6	3786	5 AAS72683	Aas72683 DNA encod
23	411	19.4	2559	9 ADD15216	Add15216 Human ser
24	410	19.3	3098	5 ABV25148	Abv25148 Human pro
25	407.5	19.2	2934	6 ABT99306	Abt99306 Mouse isc
26	406.5	19.1	2512	5 AAS67345	Aas67345 DNA encod
27	406	19.1	1422	6 ABK15142	Abk15142 Human hep
28	406	19.1	2872	6 ABL69589	Abi69589 Prostate
29	406	19.1	2872	6 ABT11015	Abt11015 Human bre
30	406	19.1	2872	7 ACC50184	Acc50184 Breast ca
31	406	19.1	11613	6 ABK15141	Abk15141 Human hep
32	405	19.1	3283	4 AAZ26570	Aaz26570 Human win
33	405	19.1	6458	6 ABK48984	Abk48984 Genomic D
34	404.5	19.1	1860	2 AAQ50632	Aaq50632 Brain fac
35	404.5	19.1	2832	2 AAQ50631	Aaq50631 Brain fac
36	403.5	19.0	1309	4 AAF27658	Aaf27658 DNA encod
37	403	19.0	2712	6 ABK48987	Abk48987 CDNA enco
38	403	19.0	6021	6 ABK48986	Abk48986 Genomic D
39	398	18.7	2448	9 ADD29815	Add29815 Human tum
40	395.5	18.6	2106	2 AAZ28103	Aaz28103 Freac11 g
41	391	18.4	1287	5 AAS72684	Aas72684 DNA encod
42	391	18.4	4454	4 AAK90985	Aak90985 Human dig
43	391	18.4	4454	5 AAS32020	Aas32020 Human liv
44	391	18.4	4454	6 ABN90375	Abn90375 Human liv
45	387	18.2	4450	4 AAK90986	Aak90986 Human dig

## ALIGNMENTS

RESULT 1

ABQ81557  
ID ABQ81557 standard; cDNA; 1209 BP.

XX AC ABQ81557;

XX DT 30-DEC-2002 (first entry)

XX DE Gene up-regulated in metastatic colorectal cancer.

XX DE Colorectal cancer; metastasis; differential expression; cytostatic;

XX DE diagnosis; gene therapy; vaccine; gene; se.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1..1209

XX FT /\*tag= a

XX FT /product= "Metastatic colon cancer polypeptide"

XX PN WO200268677-A2.

XX PD 06-SEP-2002.

XX PF 27-FEB-2002; 2002WO-US006001.

XX PR 27-FEB-2001; 2001US-027206P.

XX PR 02-APR-2001; 2001US-0281149P.

XX PR 17-APR-2001; 2001US-0284555P.

XX XX (ECSB-) BOS BIOTECHNOLOGY INC.

XX PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX Mack DH, Markowitz SD;  
 XX WPI; 2002-698677/75.  
 DR P-PSDB; ABP54691.  
 XX  
 XX New genes that are up- or down-regulated in colorectal cancer, useful for  
 PT diagnosing colorectal cancer in a subject, or for identifying modulators  
 PT of colorectal cancer-associated proteins and genes for treating  
 PT colorectal cancer.  
 XX  
 XX Claim 5; Page 252; 260pp; English.  
 XX  
 XX The present sequence is the nucleotide sequence of a human gene that  
 CC exhibits increased expression in metastatic colorectal cancer (MCC)  
 CC samples. The gene is up-regulated in colon cancer-derived metastases  
 CC compared to normal colon tissue. It is an example of claimed nucleic acid  
 CC molecules that are up- or down-regulated in metastatic colorectal cancer  
 CC cells. Such MCC-associated nucleic acids are useful in diagnostic and  
 CC prognostic applications, in screening applications e.g. biochips, for  
 CC identification of variant MCC-associated sequences, in informatics, for  
 CC expression of MCC-associated proteins, in drug screening assays for  
 CC identification of modulators of MCC-associated proteins, which are useful  
 CC for treating this cancer, in gene therapy, as DNA vaccines, and as  
 CC antisense or ribozyme modulators of MCC  
 XX  
 SQ Sequence 1209 BP; 150 A; 485 C; 432 G; 141 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 7,17e-69 Length: 1209  
 Score: 2123.00 Matches: 402  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-087-080-32 (1-402) x ABQ81557 (1-1209)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
 Db 1 ATGAAGTTGAGGTGTTTCCTCGCGCGCCACAGCGGGACCAAGCAGGCGAGTGACCTG 60  
 Qy 21 GlnGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40  
 Db 61 GAGGCGCGGGCGGACGACGCGCGCTCCCGCTGTGCGCGCGGAGGACGACTCCCTG 120  
 Qy 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGln 60  
 Db 121 GGCTCAGATGGGACTCGCGGCGCAAGCCCTCGCGGGCGCGCGCCAGAGATACGCAG 180  
 Qy 61 GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluAlaAlaProAlaAla 80  
 Db 181 GCGACGCGCAACGAGTGGGAGGCGGCGCGGCGGAGGAGGCGATCCCGGACGA 240  
 Qy 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100  
 Db 241 GCTGCTCAGCGGTGTGTGGGAGGCGGCGGAGCGCGGCGGCGGCGGCGGCGGCG 300  
 Qy 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProThrArgArgProLysProPro 120  
 Db 301 GGCGCGGGAGCGCGGAGGTGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 360  
 Qy 121 TyrSerTyrIleAlaLeuIleAlaMetAlaAlaArgAspSerAlaGlyArgLeuThr 140  
 Db 361 TACTCGTACATCGGCTCATCGCCATGCCATCGGACATCGGCGGCGGCGGCGCTTGACG 420  
 Qy 141 LeuAlaGluIleAenGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrThr 160  
 Db 421 CTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTTCGCGGCGACGTACAG 480  
 Qy 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180  
 Db 481 GGCTGGCGCAACTCCGTGGCGGCCACCACTTTGCTCAACGACTGCTTCGTCAAGGTGCTG 540

Qy 181 ArgAspProSerArgProTropGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200  
 Db 541 CGGACCCCTCGCGGCTTGGGCGAGGACAACTACTGATGCTCAACCCCAACAGGAG 600  
 Qy 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysSerHisArgAlaPro 220  
 Db 601 TACACCTTCGCGGACGGGGTCTTCGCGCGCGCGCAAGCGCTTCAGCCACCGCGCGCG 660  
 Qy 221 ValProAlaProGlyLeuArgProGluAlaProGlyLeuProAlaAlaProProPro 240  
 Db 661 GTCCCGCGCGCGGCTGCGGCGCGAGAGGCGCGGCGCTCCCGCGCGCGCGCGCGCC 720  
 Qy 241 AlaProAlaProAlaSerProArgMetArgSerProAlaArgGlnGluAlaAla 260  
 Db 721 GCGCGCGCGCGCGGCTCGCGCGATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCC 780  
 Qy 261 SerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280  
 Db 781 AGCCCCGCGGCAAGTTCCTCAGCTCTCTCCCATTCACAGCATCTTCGCGCAAGCCCTTC 840  
 Qy 281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro 300  
 Db 841 CGCAGCGCTGCGCTCAGGACACGCGCGCGCGGCGCGCTTCAGTGGGCGCGCGCGCGCC 900  
 Qy 301 CysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeu 320  
 Db 901 TGCCCGCGCGTCCCGCGCTTCCCGCGCGCTTCCCGCGCGCGCTGAGGGCGCTGCTG 960  
 Qy 321 ProLeuCysAlaTyrGlyAlaGlyValProAlaArgLeuGlyAlaArgGluAlaGluVal 340  
 Db 961 CGGCTCTCGCGGTACGCGCGCGGCGGAGCGCGCGCTGGGCGCGCGCGCGCGCGAGGTG 1020  
 Qy 341 ProProThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaProAlaLysPro 360  
 Db 1021 CCACCGACCGCGCGCGCTCTCTGCTTGCACCTCTCCCGCGCGCGCGCGCGCGCGCGCA 1080  
 Qy 361 LeuArgGlyProAlaAlaGlyValAlaHisLeuTyrCysProLeuArgLeuProAlaAla 380  
 Db 1081 CTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTACTGCCCCCTGCGCTGCGCGAGCC 1140  
 Qy 381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThrLeu 400  
 Db 1141 CTGAGGCGCGCTTAGTCCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1200  
 Qy 401 LeuAla 402  
 Db 1201 CTAGCT 1206

RESULT 2  
 ABT42112  
 ID ABT42112 standard; DNA; 1760 BP.  
 XX  
 AC ABT42112;  
 XX  
 DT 26-JUN-2003 (first entry)  
 XX  
 DE Toxicity modelling related rat gene SEQ ID No 1814.  
 XX  
 KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;  
 KW database; drug screening; toxicity assay; rat; ds.  
 OS Rattus norvegicus.  
 XX  
 PN WO200295000-A2.  
 XX  
 XX 28-NOV-2002.  
 PD  
 XX 22-MAY-2002; 2002WO-US016173.  
 PF  
 XX 22-MAY-2001; 2001US-0292335P.  
 PR 13-JUN-2001; 2001US-0297523P.  
 PR 19-JUN-2001; 2001US-0298925P.

```

PR 10-JUL-2001; 2001US-0303807P.
PR 10-JUL-2001; 2001US-0303808P.
PR 10-JUL-2001; 2001US-0303810P.
PR 28-AUG-2001; 2001US-0315047P.
PR 27-SEP-2001; 2001US-0324928P.
PR 22-OCT-2001; 2001US-0330462P.
PR 01-NOV-2001; 2001US-0330867P.
PR 21-NOV-2001; 2001US-0331805P.
PR 19-DEC-2001; 2001US-0336144P.
PR 16-DEC-2001; 2001US-0340873P.
PR 21-FEB-2002; 2002US-0357842P.
PR 21-FEB-2002; 2002US-0357843P.
PR 21-FEB-2002; 2002US-0357844P.
PR 15-MAR-2002; 2002US-0364134P.
PR 08-APR-2002; 2002US-0370144P.
PR 08-APR-2002; 2002US-0370206P.
PR 08-APR-2002; 2002US-0370247P.
PR 17-APR-2002; 2002US-0372794P.
PR 21-APR-2002; 2002US-0371679P.
XX
PA (GENE-) GENE LOGIC INC.
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX WPI; 2003-148464/14.
XX
PT Predicting at least one toxic effect of a compound, useful for toxicity
PT modeling, comprises preparing a gene expression profile of a tissue or
PT cell sample exposed to the compound, and comparing the gene expression
PT profile to a database.
XX
PS Example 4; Page; 446pp; English.
XX
CC The invention relates to a novel method of predicting at least one toxic
CC effect of a compound. The method comprises a gene expression profile of a
CC tissue or cell sample exposed to the compound, and comparing the gene
CC expression profile to a database comprising at least part of the data or
CC information given in the specification. The methods are useful for
CC predicting at least one toxic effect of a compound, predicting the
CC progression of a toxic effect of a compound, predicting the renal
CC toxicity of a compound, or identifying toxicity markers in tissues or
CC cells exposed to known renal toxin. The genes are useful as toxicity
CC markers in drug screening and toxicity assays, in monitoring disease or
CC physiological states, or disease progression. This polynucleotide
CC represents a rat DNA sequence relating to the toxic effect database
CC described in the specification. NOTE: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the World Intellectual Property
CC Organization
XX
SQ Sequence 1760 BP; 301 A; 615 C; 535 G; 309 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 4,05e-43 Length: 1760
Score: 1412.50 Matches: 300
Percent Similarity: 77.64% Conservative: 16
Best Local Similarity: 73.71% Mismatches: 78
Query Match: 66.53% Indels: 17
DB: 7 Gaps: 6

```

US-10-087-080-32 (1-402) x AET42112 (1-1760)

```

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
DB 207 ATGAAATTGGAGTG-TTCGTCCACGCGCAGC-CACGGGGACAAAATGGCAGCGATCTG 264
QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
DB 265 GAGGGGGCGGCGAGCGAGCGAGTGCATCTCCGCGCGGTGTGACATCTCTTA 324
QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyAlaArgAspThr 59
DB 325 GGCTCAGACGGGACTGTGCAGCCAAACAGCCCGGGCGGCGGCGGCGGCGGATCTG 384

```

```

QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIlePro 78
DB 385 GAAGGTGGCGGGCGAGAGGAATTCGAGTGGCGGGCGAGCACCCACAGAC----- 435
QY 79 AlaAlaAlaAlaAlaValValAlaGluGlyAla-----GluAlaGlyAlaAlaGly 96
DB 436 -----GATCCCGAGGTGACCGATGCGACGAGCAACGACGCGCTCCCGGTGGGG 483
QY 97 ProGlyAlaGlyGlyAlaGlySerGlyGluGlyAlaAlaArgSerLysProTyrThrArgArg 116
DB 484 CCGTGGCGGGCAGCGTGGCGGGGTGAGGGCGGCGGCGAGCAAGCGGTACACGGGGCGG 543
QY 117 ProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGly 136
DB 544 CCAAGCCGCCCTACTCTCTACATCGCACTCATCGCCATGGCCATCGGCACCTCCGCGGC 603
QY 137 GlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArg 156
DB 604 GGACGGCTCAGCTGGCCGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTCGG 663
QY 157 GlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPhe 176
DB 664 GGCAGCTACACGGGTGGCGCAACTCCGTGCGCCACACCTCTCGCTCAACGACTGTTTC 723
QY 177 ValIleValIleArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsn 196
DB 724 GTCAGGTGCTGGCGACCCCTCGCGCCCTGGCGGCAAGGACAATTACTGGATGCTCAAC 783
QY 197 ProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgArgLysArgLeuSer 216
DB 784 CCCAACAGCGAATACACCTTCGCGAGGGGTCTTCGCGCGCGCGCGCAAGCGCTCAGC 843
QY 217 HisArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla 236
DB 844 CACCGGACCAACAGTCTCCGCATCGGGCTACGGCGGAGGAGCCACCCGAGACTCGGGG 903
QY 237 AlaProProAlaProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGln 256
DB 904 ACCCGCAGCGCGGCGCCACCGCGCTCTCCCCAATCGCGCGCTCGCGCGCTCGCCAG 963
QY 257 GluGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeu 276
DB 964 GAGGAGGGTCCAGCGCGCGAGAGTTCTCCAGCTCTTCGCCATCGACAGCATCTTC 1023
QY 277 ArgLysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrp 296
DB 1024 AGCAAGCCGTTTCGACGCGCGCGCA-CGGCACCGCGCTCTGGGGGTGCAGCTACCTGG 1082
QY 297 GlyAlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaProCys 316
DB 1083 AGCGTGTCTCCCTGCGCGCGCTATCCGCGCTCTTCCCGCTCTCCCGCTCGTCCGCGC 1142
QY 317 ArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArg 336
DB 1143 GGTGCCCTGTCGCGCTCTGTCTTACGGCGC-GGCGAGCCACCGCTGTGGCGTGGCG 1201
QY 337 GluAlaGluValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAla 356
DB 1202 GGCGCGAGGTGCAGCGCGCGG---CCCTGTTCGTGGCGCGCTCTCCACCGCGCGC 1258
QY 357 ProAlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeu 375
DB 1259 CCAGCCAAAGCCATTTTCGAGGTCCGAGACCGCGCGCGCGCGCACTGTACTGCCCTTA 1318
QY 376 ArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyr 395
DB 1319 CGGTGCCACGCGCTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTTAC 1378
QY 396 ProValGluThrLeuLeuAla 402
DB 1379 CGGGTGCAGACGCTGCTAGCT 1399

```













DT 12-JUL-2002 (first entry)  
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 25615.  
 DE Human, cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX Homo sapiens.\*  
 OS  
 XX W0200218632-A2.  
 PN  
 XX 07-MAR-2002.  
 PD  
 XX 01-SEP-2001; 2001WC-EP010074.  
 PF  
 XX 01-SEP-2000; 2000DE-01043826.  
 PR  
 XX 05-SEP-2000; 2000DE-01044543.  
 PR  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX WPI; 2002-371829/40.  
 XX  
 XX Determining the degree of cytosine methylation in genomic DNA, useful for  
 PT diagnosis and prognosis, comprises selective hybridization of amplicons  
 PT from chemically treated DNA.  
 PT  
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
 PS  
 XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one member,  
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
 CC degree of hybridisation to both classes is determined from the label on  
 CC the amplicon. From the ratio of labels hybridised to the two classes of  
 CC oligomers, the degree of methylation is calculated. The method is used:  
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
 CC and of a wide range of diseases, e.g. cancer, disorders of the central  
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
 CC particularly by detecting mutations or single nucleotide polymorphisms  
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
 CC investigating cell differentiation. The method allows the methylation  
 CC status of many C residues to be determined simultaneously. ABQ13410-  
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
 CC for determining the degree of cytosine methylation described in the  
 CC disclosure of the invention  
 XX  
 XX SQ Sequence 658 BP; 59 A; 108 C; 222 G; 269 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 6.42e-10 Length: 658  
 Score: 483.00 Matches: 120  
 Percent Similarity: 60.89% Conservative: 3  
 Best Local Similarity: 59.41% Mismatches: 79  
 Query Match: 22.75% Indels: 0  
 DB: 6 Gaps: 0  
 US-10-087-080-32 (1-402) x ABQ39024 (1-658)  
 QY 200 GlutThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAla 219  
 DB 1 GAGTATATTTTCGCGACGGGGTTTTCGTCGTCGTCGTAAGCGTTTATGATATCGCGG 60  
 QY 220 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaProPro 239  
 DB 61 TCGGTTTTCGGGTTTCGGGTTTCGAGAGGTTTCGGGTTTTCGTCGTCGTCGTCG 120

QY 240 ProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 259  
 DB 121 TTCGGTTTCGTCGTTTCGGTTTCGTTTCGTAAGCGTTTCGTTTCGTTTCGTAAGCGC 180  
 QY 260 AlaSerProAlaGlyLysPheSerSerPheAlaLeuLeuSerPheLeuArgLysPro 279  
 DB 181 GTTAGTTTCGGGGTAAAGTTTTCGTTTCGTTTCGTAAGCGTTTTCGTAAGCGTTTTCG 240  
 QY 280 PheArgSerArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAla 299  
 DB 241 TTTTCGTAGTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCG 300  
 QY 300 ProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeu 319  
 DB 301 TTTTCGTAGTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCG 360  
 QY 320 LeuProLeuCysAlaTyrglyValaGlyGluProAlaArgLeuGlyAlaArgGluAlaGlu 339  
 DB 361 TTGTCGTTTCGCGGTACGGCGCGGGGAGTCGCGCGGGTTGGCGCGCGCGAGTCGAG 420  
 QY 340 ValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLys 359  
 DB 421 GTGTTATCGATCGCGTCGTTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCG 480  
 QY 360 ProLeuArgGlyProAlaAlaGlyValaHisLeuTyrcysProLeuArgLeuProAla 379  
 DB 481 TTATTCGAGGTTTCGCGGTTCGCGCGGGCGGTATTTGTTGTTTTCGCGTTTCGTA 540  
 QY 380 AlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrcysProValGluThr 399  
 DB 541 GTTTTGTAGCGGGTTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCG 600  
 QY 400 LeuLeu 401  
 DB 601 TTTTAA 606  
 RESULT 10  
 AA231671  
 ID AA231671 standard; DNA; 3946 BP.  
 XX  
 AC AA231671;  
 XX  
 DT 17-JAN-2000 (first entry)  
 XX  
 DE Human FKHL7 coding sequence.  
 XX  
 KW FKHL7; human; forkhead transcription factor gene; diagnosis; therapy;  
 KW congenital heart disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9952415-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 14-APR-1999; 99WO-US008159.  
 PR 15-APR-1998; 98US-0081870P.  
 PR 22-MAY-1998; 98US-00083351.  
 XX  
 PA (IOWA) UNIV IOWA RES FOUND.  
 XX  
 PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 DR WPI; 1999-620257/53.  
 DR P-PSDB; AAY43260.  
 XX  
 PT New isolated human forkhead transcription factor gene, FKHL7, used to  
 PT develop products for the diagnosis, prognosis, monitoring, prevention or  
 PT treatment of congenital heart disease.  
 XX  
 PS Claim 11; Fig 1; 98pp; English.  
 XX

CC This sequence encodes the human forkhead transcription factor gene, designated FKHL7, of the invention. FKHL7 can be used in a novel method for treating or preventing the development of a congenital heart disease (CHD) in a subject. The FKHL7 sequences can be used for diagnosis, prognosis, monitoring, prevention and treatment of CHD. They can also be used for the production of transgenic animals and drug screening.

XX  
SQ Sequence 3946 BP; 905 A; 1155 C; 1053 G; 833 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 6.04e-09 Length: 3946  
Score: 472.50 Matches: 165  
Percent Similarity: 39.07% Conservative: 44  
Best Local Similarity: 30.84% Mismatches: 145  
Query Match: 22.26% Indels: 182  
DB: 2 Gaps: 22

US-10-087-080-32 (1-402) x AAZ31671 (1-3946)

QY 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27  
DB 287 CCGAGGCGCGGAGGAGCAGCCAGCGAGCGCGGAGGAGGCGGAGCGAGCGCGGAGC 346  
QY 28 Ala-----ProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43  
DB 347 GCACAGCGCAGCGGCGCGGACCA---GCTCGCGCGGCGCGGACTCGGACTCGGCGGCC 403  
QY 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAspGly 63  
DB 404 GCGCGGCGCGCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463  
QY 64 GluGlnSerAlaGlyGlyGlyProGlyAlaGluGlnAlaAlaAla-----81  
DB 464 GAGCGG-----GGGCCA---TGCAAGGCGCGTACTCGGTCTCCAGCGCCCAACT 508  
QY 82 -----AlaAlaAlaValAlaGluGlyAlaGluAlaGly 93  
DB 509 CCCTGGGAGTGTGGCTACTCGCGGCGGAGAGAGTACTACCGC-GGCGGCGCGCG 567  
QY 94 AlaAlaGlyProGly-----AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyr 113  
DB 568 GCGGCGGCGGCGGCGTACACCGCCATCGCGCCCGCCATGACGCGTGTACTCGCACCCCTGCG 627  
QY 99 -----AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyr 113  
DB 628 CACGCGGAGAGTACCGCGGCGGATGCGCGCGCCCGCTACGCG-----CCCTAC 675  
QY 114 ThrArgArgPro-----LysProProTyrSerTyrIleAlaLeuIle 127  
DB 676 ACGCGCAGCGCGCAGCCCAAGGACATGTTGAAGCGCCCTATAGCTACATCGCGTCTATC 735  
QY 128 AlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyr 147  
DB 736 ACCATGGCCATCCAGAGCGCCCGGACAAAGAGATCAACCTGGAACGCGATCTACCACTTC 795  
QY 148 LeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSerValArg 167  
DB 796 ATCATGGACCGCTTCCCTTCTACCGGACACACAGCAGGCGCTCGCAGACAGCATCCGC 855  
QY 168 HisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArgProTyr 187  
DB 856 CACAACCTCTCGCTCAACAGAGTGTCTGTCAGGTGCGCGGACGACAAAGAGCGCC---912  
QY 188 GlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyVal 207  
DB 913 GCGAAGGCGAGTACTGACCGCTGACCGGACTCTTACACATGTTTCGAGACGCGGAGC 972  
QY 208 PheArgArgArgLysArgLysSerHis-----217  
DB 973 TTCTGCGGCGCGCGCGCTTCAAGAAAGAGGAGCGGTGAAGGACAAAGGAGGAGAG 1032  
QY 218 -----ArgAlaProValProAlaProGly 225

DB 1033 GACAGGCTGCACCTCAAGAGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1089  
QY 226 LeuArgProGluGluAlaProGlyLeuProAlaAlaPro-----238  
DB 1090 -----CCGAGCAGCATC 1143  
QY 239 -----ProProAlaProAlaAlaProAla 246  
DB 1144 CAGGACATCAAGACCGAGAACGGTAGTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1203  
QY 247 Ser-----ProArgMetArgSerPro-----253  
DB 1204 GCGCGCTCGGCGAGCAGC 1263  
QY 254 ---AlaArgGlnGluGluArgAlaSerProAlaGlyLysPhe-----266  
DB 1264 AGCAGGAGCTGTCCAGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1323  
QY 266 -----266  
DB 1324 AGCCTGGACGGTGGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1383  
QY 267 ---SerSerSerPheAlaIleAspSerIleLeu-ArgLys-----278  
DB 1384 CATAGCCAGGGCTTCAGCGTGGACACATCATGCTCGTGGGGGGTGGCGCGCGCGCGCAGC 1443  
QY 279 -----ProPheArgSerArgArgLeuArgAspThrAlaProGl 291  
DB 1444 GCGCGCGCGAGCTCAGCTCCGCGCTTCTGGCTCGCGCGCGCGCGCGCGCGCGCGCGG 1503  
QY 291 YThrThrLeuGlnTyrGlyAlaAlaProCysProProLeuProAlaPheProAlaLeu 311  
DB 1504 ATCGCACCCCGCGCGCTCGCGCGCTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1563  
QY 311 uProAlaAlaProCysArgAlaLeuProLeuCysAlaTyrGlyAlaGlyGluProAl 331  
DB 1564 CCCTGCGCGCGAGCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 1623  
QY 331 aArgLeuGlyAlaArgGluAlaGluValProThrAla-----ProProLeuLeu 348  
DB 1624 GCGG---GGGCTG 1680  
QY 348 uLeuAlaProLeu-----ProAlaAlaAlaProAla 359  
DB 1681 TAGCTCG 1740  
QY 359 sProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysPro-----Le 375  
DB 1741 GCGGTGGACAAACCGCG-----TGCCGCGACTACTCTCTGTGCT 1776  
QY 375 uArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgPro 389  
DB 1777 CCGGTCCACGAGCGCA-----CCTCGTGTGCTCC 1804  
RESULT 11  
AAZ38079  
ID AAZ38079 standard; DNA; 3946 BP.  
XX  
AC AAZ38079;  
XX  
DT 22-FEB-2000 (first entry)  
XX  
DE Human forkhead transcription factor gene, FKHL7 DNA sequence.  
XX  
KW Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
XX  
KW transgenic animal; drug screening; ss.  
XX  
OS Homo sapiens.  
XX  
FN W0953060-A2.  
XX  
PD 21-OCT-1999.  
XX

PF 14-APR-1999; 99WO-US008148.  
 PR 15-APR-1998; 98US-0081870P.  
 PR 22-MAY-1998; 98US-0083352.  
 XX (TOWA) UNIV IOWA RES FOUND.  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 PI WPI; 1999-620429/53.  
 DR P-PSDB; AAY55721.  
 XX New isolated human forkhead transcription factor gene, FKHL7, used to,  
 PT e.g. develop products for the diagnosis.  
 XX Claim 1; Fig 1; 99pp; English.  
 XX The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. The present sequence represents  
 CC the DNA sequence of the FKHL7 gene  
 XX  
 SQ Sequence 3946 BP; 905 A; 1155 C; 1053 G; 833 T; 0 U; 0 Other;  
  
 Alignment Scores:  
 Pred. NO.: 6.04e-09 Length: 3946  
 Score: 472.50 Matches: 165  
 Percent Similarity: 39.07% Conservative: 44  
 Best Local Similarity: 30.84% Mismatches: 145  
 Query Match: 22.26% Indels: 182  
 DB: 2 Gaps: 22  
  
 US-10-087-080-32 (1-402) x AA238079 (1-3946)  
  
 QY 8 ProAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27  
 DB 287 CCGAGCGCGGAGGAGCGAGCCAGCCAGCGCGGAGCGGCGGCGGCGGCGGCGGAC 346  
 QY 28 Ala-----ProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43  
 DB 347 GCACAGCGCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 403  
 QY 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaAlaArgAspThrGlnGlyAspGly 63  
 DB 404 GCGCGCGCGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 463  
 QY 64 GluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaLeuProAlaAlaAla----- 81  
 DB 464 GAGCGG-----GGGCGCA---TGCAGCGCGCGCTACTCCGTGTCCAGCGCGCGCAACT 508  
 QY 82 -----AlaAlaAlaValValAlaGluGlyAlaGluAlaGly 93  
 DB 509 CCCTGGGAGTGGTCCCTACCTCGCGCGGCGGAGAGCTACTACCGC-GCGGCGGCGCG 567  
 QY 94 AlaAlaGlyProGly----- 98  
 DB 568 GCGGCGGCGGCGGCTACACCGCCATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 627  
 QY 99 -----AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyr 113  
 DB 628 CACGCGGAGAGTACCGCGGCGGCGGATGCGCGCGCGCTACGGG-----CCCTAC 675  
 QY 114 ThrArgArgPro-----LysProProTyrSerTyrIleAlaLeuIle 127  
 DB 676 AGCGCGCAGCGCGAGCGAGCATGTGTGAAGCGCGCTATAGCTACATCGCGCTCATC 735  
 QY 128 AlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrIleAlaGluIleAsnGluTyr 147  
 DB 736 ACCATGGCATCCAGAACGCGCGGACAAAGAGATACCCCTGAACGCGCATCTACCATTC 795  
 QY 148 LeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSerValArg 167

RESULT 12  
 AA228095

DB 796 ATCATGGACCGCTTCCCTTCTACCGGCAACAAGCAGGCGCTGGCAGAACGATCCGC 855  
 QY 168 HisAsnLeuSerLeuAsnAspCysPheValIysValLeuArgAspProSerArgProTyr 187  
 DB 856 CACAACCTCTCGCTCAACGAGTCTTCTCGTCAAGTTCGCGCGCAGCAAGAAGCGC--- 912  
 QY 188 GlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyVal 207  
 DB 913 GCGAAGGCGCAGCTACTGTGACGCTGACCGCGGACTCTTACACATGTTTCGAGAACGCGAGC 972  
 QY 208 PheArgArgArgLysArgLysArgLysHis----- 217  
 DB 973 TTCCTGGCGCGCGCGCGCTTCAAGAGAAGAGCGGTGAAGGACAAAGAGAGAGAAG 1032  
 QY 218 -----ArgAlaProValProAlaProGly 225  
 DB 1033 GACAGGCTGCACCTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC--- 1089  
 QY 226 LeuArgProGluGluAlaProGlyLeuProAlaAlaPro----- 238  
 DB 1090 -----CCGGAGCAGCGCGCGCAACCGCGCGCGTCCGACGCGCGCGCGCGCGCATC 1143  
 QY 239 -----ProProAlaProAlaAlaProAla 246  
 DB 1144 CAGGACATCAAGACCGAGACCGTACGTGCCCTCGCGCGCGCGCGCGCGCGCGCGCGC 1203  
 QY 247 Ser-----ProArgMetArgSerPro----- 253  
 DB 1204 GCGCGCGTGGCGAGCGCGCGCGCGCGCGCGCGTGTCCAGATCGAGAGCGCGCGCGCGCAGC 1263  
 QY 254 ---AlaArgGlnGluGluArgAlaSerProAlaGlyLysPhe----- 266  
 DB 1264 AGCAGCAGCCTGTCCAGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1323  
 QY 266 ----- 266  
 DB 1324 AGCTTGGACGTGGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 1383  
 QY 267 ---SerSerSerPheAlaIleAspSerIleLeu-ArgLys----- 278  
 DB 1384 CATAGCCAGGGCTTCAGCGTGCAGCAACATCATGACGTGCTCGCGGGGTGCGCGCAGAGC 1443  
 QY 279 -----ProPheArgSerArgLeuArgAspThrAlaProGly 291  
 DB 1444 GCGCGCGCGGAGCTCAGTCCGCGCTTCTGCGCTCGCGCGCGCGCGCGCGCGCGCGCGG 1503  
 QY 291 YThrThrLeuGlnTyrGlyAlaAlaProCysProProLeuProAlaPheProAlaLeuLe 311  
 DB 1504 ATCGACCCCGCTGCGCGCTGCGCGCTACTCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 1563  
 QY 311 uProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAl 331  
 DB 1564 CCTGTGAGCCAGACTCCAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 1623  
 QY 331 aArgLeuGlyAlaArgGluAlaGluValProThrAla-----ProProLeuLe 348  
 DB 1624 GCGG---GGGCTG 1680  
 QY 348 uLeuAlaProLeu-----ProAlaAlaAlaProAlaLys 359  
 DB 1681 TAGCGCGCGCGCGAGCTG 1740  
 QY 359 sProLeuArgGlyProAlaAlaAlaGlyGlyAlaHisLeuTyrCysPro-----Le 375  
 DB 1741 GCGGTGGCAACCCCG-----TGCGCGACTACTCTCTCGCT 1776  
 QY 375 uArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgPro 389  
 DB 1777 CCGGTCAACGAGCAGCA-----GCTCGTGTGCGC 1804

ID AAZ28095 standard; cDNA; 1976 BP.  
 AC AAZ28095;  
 XX  
 DT 31-JAN-2000 (first entry)  
 DE Human FREAC3 protein encoding cDNA.  
 XX  
 DE FREAC3 gene; eye disease; eye developmental defect; therapy; mutation;  
 KW glaucoma; anterior segment dysgenesis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 159..1820  
 FT /\*tag= a  
 FT /product= "FREAC3"  
 FT 403  
 FT mutation  
 FT /\*tag= b  
 FT /note= "a G to C transversion mutation at this position  
 FT can result in a Ser82Thr mutation in helix 1 of the  
 FT FREAC3 forkhead domain"  
 FT 419  
 FT mutation  
 FT /\*tag= c  
 FT /note= "a G to C transversion mutation at this position  
 FT can result in a Ile87Met mutation in helix 1 of the  
 FT FREAC3 forkhead domain"  
 XX  
 PN WO954493-A2.  
 XX  
 XX 28-OCT-1999.  
 PD  
 PF 16-APR-1999; 99WO-IB001024.  
 XX  
 PR 17-APR-1998; 98US-0082206P.  
 PR 08-MAY-1998; 98US-0084784P.  
 XX  
 XX (UYAL-) UNIV ALBERTA.  
 PA  
 XX Walter MA, Jordan T, Raymond V;  
 PI  
 XX WPI; 1999-634007/54.  
 DR P-PSDB; AAY41277.  
 DR  
 XX  
 PT Use of mutations in the FREAC3 gene for detecting predisposition to eye  
 PT diseases or developmental defects and for developing products for  
 PT prevention and therapy.  
 PT  
 PS Claim 50; Fig 2; 65pp; English.  
 XX  
 CC The invention relates to the use of mutations in the FREAC3 gene for  
 CC detecting predisposition to diseases of the eye or developmental defects  
 CC and for developing products for prevention and therapy. The method  
 CC comprises analyzing nucleic acid of the mammal to determine whether the  
 CC nucleic acid contains a mutation in a FREAC3 gene, where the presence of  
 CC the mutation is an indication that the mammal has an increased likelihood  
 CC of developing a disease of the eye. The methods can be used for the  
 CC diagnosis of increased likelihood of developing a disease of the eye,  
 CC e.g. glaucoma, or a developmental defect. The methods can also be used  
 CC for the prevention and treatment of diseases of the eye. Transgenic  
 CC animals containing the mutated FREAC3 gene can be used to screen for  
 CC therapeutic compounds that modulate anterior segment dysgenesis or the  
 CC onset or progression of glaucoma via a FREAC3-dependent or FREAC3-  
 CC affected pathway. The present sequence represents the cDNA encoding the  
 CC FREAC3 polypeptide  
 XX  
 SQ Sequence 1976 BP; 362 A; 736 C; 658 G; 220 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 5,86e-09 Length: 1976  
 Score: 466.50 Matches: 164  
 Percent Similarity: 39.24% Conservative: 42  
 Best Local Similarity: 31.24% Mismatches: 133

Query Match: 21.97% Indels: 187  
 DB: 2 Gaps: 22  
 US-10-087-080-32 (1-402) x AAZ28095 (1-1976)  
 QY 22 GlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu-Gl 41  
 DB 6 GGGAGAGGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 65  
 QY 41 YSerAspGlyAspCysAlaAla-----LysProSerAlaGlyGlyAl 56  
 DB 66 CCGGAGCTCGGACTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125  
 QY 56 aArgAspThrGlnGlyAspGlyGlnSerAlaGlyGlyGlyProGlyAlaGluAl 76  
 DB 126 GCGGGCGGCGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 170  
 QY 76 aileProAlaAlaAla-----AlaAlaAlaValva 86  
 DB 171 TACTCGGTGTCCAGAGCCCAACTCCCTGGGAGTGGTCCCTTACCTCGGCGGAGCAGC 230  
 QY 86 lAlaGluGlyAlaGluAlaGlyAlaGlyProGly----- 98  
 DB 231 TACTACGCG-GCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 289  
 QY 99 -----AlaGlyGlyAlaGlySerGlyGl 106  
 DB 290 GAGCGTGTACTCGACCCCTCGCACCGCGAGCAGTACCCTGGGCGGCGGCGGCG 349  
 QY 106 uGlyAlaArgSerLysProTyrThrArgArgPro-----LysProPr 120  
 DB 350 CGGG-----CCCTACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 397  
 QY 120 oTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyLeuTh 140  
 DB 398 CTATAGCTACATCGCGCTCATCACCATGGCCATCCAGACGCCCGCGGACAGATCAC 457  
 QY 140 rLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrTh 160  
 DB 458 CCTGACGCGCATCTACCGATTCATCATGGACCCCTTCCCTTCTACGCGGACACAGCA 517  
 QY 160 rGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheVallysValle 180  
 DB 518 GGGCTGGCAAGACAGATCCGCCACACCTCTCGCTCAAGAGTGTCTCGTCAAGTGCC 577  
 QY 180 uArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGl 200  
 DB 578 GCGCGACGACACAGAGCGCG--GGCAAGGCGAGCTACTGGACGCTGGACCGGACTCCTA 634  
 QY 200 uTyrThrPheAlaAspGlyValPheArgArgArgLysArgLysArgLysSerHisArg----- 218  
 DB 635 CAACATGTTTCGAGAACGCGGAGCTTCCTGGGCGGCGGCGGCGGCGGCGGCGG 694  
 QY 218 ----- 218  
 DB 695 GTTGAAGGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 754  
 QY 219 -AlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaPr 238  
 DB 755 CAGCCCCCGCGCGCGCGCG-----CGGAGAGCGCGCGCGCGCGCGCGCGCG 805  
 QY 238 o-----Pr 239  
 DB 806 GCAGCG 865  
 QY 239 oProAlaProAlaAlaProAlaSer-----ProAr 249  
 DB 866 GCGCCAGCG 925  
 QY 249 gMetArgSerPro-----AlaArgGlnGluAlaArgAlaSerProAlaGlyLy 265  
 DB 926 GATCGAGAGCG 985

```
Qy 265 sPhe----- 266
Ps 986 CTGCGCTCGCGCGCGCTAGCTGAGCTGAGCGTGGATTCGCGCGCGCGCGCGCGCG 1045
Cc 267 -----SerSerSerPheAlaIleAspSerIleLeu-ArgL 278
Cc 1046 GCCTCGCGCGCGCGCGCGCAGCATAGCCAGGGCTTCAGCGTGGACAACATCATGACGTC 1105
Cc 278 ys-----ProPheArgSerArgA 284
Cc 1106 GCTCGCGGGGTGCGCGCAGAGCGCGCGCGGAGCTCAGCTCGGCTTCTGGCTCGGC 1165
Cc 284 rgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaProCysProProL 304
Cc 1166 GGCGCGCTCTCGCGCGCGGAGTCGACCCCGCTGGCGCTCGGCGCTACTCGCGCGG 1225
Cc 304 euProAlaPheProAlaLeuLeuProAlaAla-----ProCysArgAlaLeuLeuProL 322
Cc 1226 CCAGAGCTCCCTACAGCTCCCTCGAGCCAGACCTCCAGCGCGGCGAGCTCGGCGG 1285
Cc 322 euCysAlaTyrglyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProP 342
Cc 1286 CGCGCGCGCGCGCGCGGCGG-----CGGGGGCGCGCGCGCGCGCGGACCTA 1336
Cc 342 roThrAla-----ProProLeuLeuAlaProLeu----- 352
Cc 1337 CCACTGCAACCTGCAAGCCATGAGCCTGTAGCGCGCGCGAGCGCGGCGGCGGCACTTGA 1396
Cc 353 -----ProAlaAlaProAlaLysProLeuArgGlyProAlaAlaAlaGlyAlaH 370
Cc 1397 GGGCGCGCGCGCGCGCGGCGCTCGCGCTGGAGCAGCCCG----- 1440
Cc 370 isLeuTyrcysPro-----LeuArgLeuProAlaAlaLeuGlnAlaLeuV 386
Cc 1441 -----TGCCGACTACTCTCTGCTCGCTCGGTCCAGCAGCA-----G 1477
Qy 386 alArgArgPro 389
Cc 1478 CTCGTGTCCTCC 1488

RESULT 13
ABZ34783
ID ABZ34783 standard; cDNA; 2187 BP.
XX AC ABZ34783;
XX DT 04-FEB-2003 (first entry)
XX DE Coding sequence SEQ ID 141, upregulated in osteogenesis.
XX KW Osteopathic; osteogenesis modulator; gene therapy; osteogenesis;
XX KW osteoporosis; bone disease; downregulator; human; singal transduction;
XX KW ss.
XX OS Homo sapiens.
XX PN WO20021745-A2.
XX PD 17-OCT-2002.
XX PF 05-APR-2002; 2002WO-IB002211.
XX PR 05-APR-2001; 2001US-0281400P.
XX PA (AVET ) AVENTIS PHARMA SA.
XX PI Garcia T, Roman Roman S, Baron R, Call K, Theilhaber J;
XX PI Connolly T, Jackson A, Bushnell SE, Rawadi G;
XX DR WPI; 2003-058567/05.
XX PT Novel isolated nucleic acid upregulated/downregulated in osteogenesis,
XX PT useful for bone disease therapy in subject.
```

```
XX Claim 26; Page 153-154; 237pp; English.
XX The present invention relates to novel nucleotide sequences, which are
Cc differentially expressed in models of osteogenesis upon being put in
Cc contact with a stimulator of osteogenesis. The present sequence is one
Cc such sequence. This sequence can be used for diagnosing osteoporosis/bone
Cc disease in a patient, promoting osteogenesis and/or preventing
Cc osteoporosis/bone disease. The present sequence encodes a protein
Cc involved in intracellular signal transduction
XX SQ Sequence 2187 BP; 411 A; 736 C; 615 G; 425 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.33e-09 Length: 2187
Score: 466.50 Matches: 157
Percent Similarity: 45.35% Conservative: 48
Best Local Similarity: 34.73% Mismatches: 159
Query Match: 21.97% Indels: 90
DB: 7 Gaps: 18

US-10-087-080-32 (1-402) x ABZ34783 (1-2187)
Qy 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27
Db 79 CGCGCTCGCGCGCGCGCTCGCTCCCGGTCCAGATGACACCGAGCGGGG---CCG 135
Qy 28 AlaProSerProLeuSerAlaAlaGlyAspAspSerLeuGlySer---AspGlyAspCys 46
Db 136 CGCGCGCGCGCTCGCGCGCGGTGCGCGCGGTCCCGCGCGCTCCAGCGCGCGCTG 195
Qy 47 AlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGlnGlyAspGlyGluGlnSer 66
Db 196 ATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 255
Qy 67 AlaGlyGlyGlyProGlyAlaGluAlaLarLeProAlaAlaAlaAlaAlaValVal 86
Db 255 TCCTCGTCTGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCTCGCTCT 315
Qy 87 AlaGluGlyAlaGluAlaAlaGlyAlaGlyProGlyAlaGlyGlyAlaGlySerGlyGlu 106
Db 316 GCGCGCTCGCTCGCTCGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 375
Qy 107 GlyAlaArgSerLysProTyrrArgArgPro---LysProProTyrrSerTyrrIleAla 125
Db 376 GCGAAGAGCGAGCTCGGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 435
Qy 126 LeuIleAlaMetAlaAlaArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsn 145
Db 436 CTCATCGTCATGCGCATCCAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495
Qy 146 GluTyrrLeuMetGlyLysPheProPhePheArgGlySerTyrrThrGlyTrpArgAsnSer 165
Db 496 CAGTTCCTGCGCGCGCTTCCTCTTCCGCGCGCTTCCAGGGCTGGAGAACTCG 555
Qy 166 ValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArg 185
Db 556 GTGCGCCACAATCTCTCGCTCAACGAGTCTTCATCAGCTGCTTAAGGGCTCGGCGG 615
Qy 186 ProTrpGlyLysAspAsnTyrrTrpMetLeuAsnProAsnSerGluTyrrThrPheAlaAsp 205
Db 616 CCC---GGCAAGCGCCACTTACTGGACCATCGACCGCGCGCGCGCGCGCGCGCG 672
Qy 206 GlyValPheArgArgArgLysArgLysSerHisArgAlaPro----- 220
Db 673 GGTCTGTTCCCGCGCGGCG---GCGCGGCTTCAGCGGAGTCCAGCGCTCAAGCCCAT 731
Qy 221 ValPro-----AlaProGlyLeuArg 227
Db 732 GTACCAACCGCTGTGAGCGGCTTGGGCTTGGGCGTCTGCTGCTGCTGCTGCTGCTG 791
Qy 228 ProGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaAlaProAlaSer 247
```

Db 792 -----CTTCAGGGGCCCCCGTCCGGCGCGCTCGGCTGCCAGGCCA 833  
 Qy ProArgMetArgSerPro-----AlaArgGln 256  
 Db 834 GGGCGGCTACGGCGGCTCGACATGATGCCCGGGCTACGACGCGCGGGCGGCC 893  
 Qy 257 GluGluArgAla-----SerProAlaGlyLysPheSer 268  
 Db 894 CAGCCAGGGGACCTCCACCACCACCACCACCGTCCGCGCATGTGCCCAACC 953  
 Qy 269 SerPheAlaAlaAspSerIleLeu-----ArgLysProPhe 280  
 Db 954 GGGTTCACCTACATGCGCAGCTGCGCGTCCCGGGAGCCCGGGGCGTCCGTGCGGC 1013  
 Qy 281 ArgSerArgArgLeuArg-AspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPr 300  
 Db 1014 CCGGGCGGCGCGCGGGGACTACGGCGGACAGCAGC-----AGCAGGCC 1051  
 Qy 300 cCysProProLeuProAlaPhe-----ProAlaLeuLeuPr 312  
 Db 1062 GTACCCCTCGTCCCGGCCATGCGGAGCGCCATCGAATGCCACTCGCCCTACAGAGCC 1121  
 Qy 312 alaAla-----ProCysArgAlaLeuLeuPro-----LeuCysAlaTy 325  
 Db 1122 TCGCGCGCACTGGAGCTCGCCTGG-CGCTCGCCTTACCTCAAGCAGCGCCTGCTCGCTGA 1180  
 Qy 325 rGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAlaPr 345  
 Db 1181 CCGCCAGCAGCAACCCCGCGCTGGCAGG-----CTGCACTCCAGCA 1225  
 Qy 345 oProLeuLeuAlaProLeuProAlaAlaProAlaProAlaProLeuArgGlyProAl 365  
 Db 1226 TGTCTCTCTACTCGCTGGAGCAGAGCTACTTGCACCAAGAGCGCTCG---CGAGGACCTCT 1282  
 Qy 365 alaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaAlaLe 385  
 Db 1283 CAGTGGGACTGCC-----CCGTACAGCATCACTCTACTCGAGTGT 1324  
 Qy 385 u---ValArgArgProGlyProHisLeuSerTyr 395  
 Db 1325 GTGACAGAAAGATTTCGTCTCAACTTCAATGG 1358  
 RESULT 14  
 ABX76288  
 ID ABX76288 standard; DNA; 3482 BP.  
 XX AC ABX76288;  
 XX DT 02-APR-2003 (first entry)  
 XX DE Lung cancer-associated polynucleotide #152.  
 XX KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;  
 XX KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
 XX OS Unidentified.  
 XX PN WO200285443-A2.  
 XX PD 31-OCT-2002.  
 XX PF 18-APR-2002; 2002WO-US012476.  
 XX PR 18-APR-2001; 2001US-0284770P.  
 PR 10-MAY-2001; 2001US-0290492P.  
 PR 09-NOV-2001; 2001US-0339245P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 29-NOV-2001; 2001US-0334370P.  
 PR 12-APR-2002; 2002US-0372246P.  
 XX

PA (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX Aziz N, Murray R;  
 XX WPI; 2003-093161/08.  
 XX P-PSDB; ABUS6559.  
 XX  
 PT Detecting a lung cancer-associated transcript in a cell from a patient  
 PT for treating lung cancer, by contacting a biological sample from the  
 PT patient with a polynucleotide that exhibits increased or decreased  
 PT expression in lung cancer.  
 XX  
 PS Claim 22; Page 303-304; 453pp; English.  
 XX  
 CC The invention relates to a method for detecting a lung cancer-associated  
 CC transcript in a cell from a patient, comprising contacting a biological  
 CC sample from the patient with a polynucleotide that selectively hybridizes  
 CC to a sequence that is at least 80 % identical to a gene that exhibits  
 CC increased or decreased expression in lung cancer samples. Lung cancer-  
 CC associated polynucleotides and polypeptides are used for identifying a  
 CC compound that modulates a lung cancer-associated polypeptide, for  
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung  
 CC cancer in a patient and for treating a mammal having lung cancer by  
 CC administering a modulatory compound identified. The methods are useful  
 CC for treating lung cancer, such as small cell lung cancer, non-small cell  
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful  
 CC for diagnostic purposes and as targets for screening for therapeutic  
 CC compounds that modulate lung cancer, such as antibodies. Sequences  
 CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the  
 CC invention  
 XX  
 SQ Sequence 3482 BP; 716 A; 1137 C; 981 G; 647 T; 0 U; 1 Other;  
 Alignment Scores: Length: 3482  
 Pred. No.: 1.69e-08 Matches: 157  
 Score: 459.00 Conservative: 34  
 Percent Similarity: 48.72% Mismatches: 142  
 Best Local Similarity: 40.05% Indels: 63  
 Query Match: 21.62% Gaps: 18  
 DB: 7  
 US-10-087-080-32 (1-402) x ABX76288 (1-3482)  
 Qy 30 SerProLeuSerAlaAlaGly-----AspAspSerLeuGlySer--- 42  
 Db 504 GCGCCAGCTGGGCGCGCTCGGCTCTCCGGCCAGCCCGCGAGCATCCCTGAGCTCTCG 563  
 Qy 43 AspGlyAspCysAlaAlaLysProSerAlaGlyGly-----GlyAlaArgAspThrGln 60  
 Db 564 CAGAAGGGCGAGCGTCCGTTCGGGGAGCGCCAGGCCCGCCCGCCCGCCGACAGCCGC 623  
 Qy 61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGly-----AlaGluGluAlaIle 77  
 Db 624 GGGGATCCAGAGCCCGGGGTGGCGAGCGCCCGCCCATGACTGCGGAGAGCGGGCGCCG 683  
 Qy 78 ProAlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGly-----AlaAla 95  
 Db 684 GCCCGCGCAGCGGAGGTGCTGCTACCGTGAAGGAA--GAGCGCGGAGAGCGGAGCA 741  
 Qy 96 GlyProGlyAlaGlyGlyAlaGlySerGlyGluGlyAla-----ArgSerLysPro 112  
 Db 742 GGGCGCGGTCCAGGGGAGGCCACGGCCCGGGCGGGCGGGCGGCCCAAGCGC 801  
 Qy 113 TyrThrArgArgProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArg 132  
 Db 802 CCCCTGAGCGCGGAGCGCCCTACGCTACATCGCGCTCATCGCCATGCGCATCGCG 861  
 Qy 133 AspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPhe 152  
 Db 862 CAGCGCGCGAGCGCGCTCACGCTGGGCGGATCTACAGTTCATCAGCGAGCGCTTC 921



```
Db 337 CGCCACAACTCTCTCGCTCAAGAGTGGTTTCATCAAGCTGCTAAGGCCCTCGGGCGGCC 396
Qy 187 TrpGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGly 206
Db 397 ---GGCAAGGGCCACTACTGACCATCGACCCGCCAGCGAGTTTCATGTTTCGAGGAGGC 453
Qy 207 ValPheArgArgArgLysArgLeuSerHisArgAlaPro-----Val 221
Db 454 TCGTTCGCGCCCGCGCC- GCGCGGCTTCAGGCGGAAGTGCAGCGCTCAAGCCCATGTA 512
Qy 222 Pro-----AlaProGlyLeuArgPro 228
Db 513 CCACCGCTGTGTAGCGGCTTGGGCTTCGGGGCTCGCTGCTGCCCGCCAGGGCTTCGA-- 569
Qy 229 GluGluAlaProGlyLeuProAlaAlaProProAlaProAlaAlaProAlaSerPro 248
Db 570 -----CITCCAGGCGCCCGCTCGCGCGCGCTCGGCTGCCACAGCCAGGG 614
Qy 249 ArgMetArgSerPro-----AlaArgGlnGlu 257
Db 615 CGGCTACGGCGGCTCGACATGATGCCCGGGGTACGACGCCGGCGGGCGGCCCCAG 674
Qy 258 GluArgAla-----SerProAlaGlyLysPheSerSerSer 269
Db 675 CCAGCGCGCACCCCTCACCACCACCACCACCACCATGTCGCGCACATGTCGCCCAACCGGG 734
Qy 270 PheAlaIleAspSerIleLeu-----ArgLysProPheArg 281
Db 735 TTCACCTACATGGCCAGCTCCCGGTGCGCGGGACCGGGGGCTCGGTGCGGCGG 794
Qy 282 SerArgArgLeuArg-AspThrAlaProGlyThrLeuGlnTrpGlyAlaAlaProCy 301
Db 795 GGGCGGCGGGCGCGGCGGACTACGGCGCGGACAGCAGC-----AGCAGCCCGGT 842
Qy 301 sProProLeuProAlaPhe-----ProAlaLeuLeuProAl 313
Db 843 ACCCTCGTCCCGGCCATGGCGAGGCCATCGAATGCCACTCGGCCCTACAGAGCCCTGC 902
Qy 313 aAla-----ProCysArgAlaLeuLeuPro-----LeuCysAlaTyrGl 326
Db 903 GCGCGACTGGAGCTCGCCTGG- CGCCTCGCCTTACCTCAAGCAGCCGCTGCCCTGACGC 961
Qy 326 yAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAlaProPr 346
Db 962 CCAGCAGCAACCCCGCGCCTCGGCAGG-----CCTGCACCTCCAGCATGT 1006
Qy 346 oLeuLeuAlaAlaProLeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAl 366
Db 1007 CCTCTACTCGCTGGAGCAGAGTACTTGACACAGACGCTCG---CGAGGACCTCTCAG 1063
Qy 366 aGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaLeu-- 385
Db 1064 TGGGACTGCC-----CCGTATCAGCATCACTTACTTACTTACTTACTTACTTACTT 1105
Qy 386 -ValArgArgProGlyProHisLeuSerTyr 395
Db 1106 ACAGAAAGATTTTGTCTCTCAACTTCAATGG 1136
```

Search completed: March 31, 2004, 06:46:05  
Job time : 467 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 31, 2004, 05:56:45 ; Search time 100 Seconds  
(without alignments)  
2230.903 Million cell updates/sec

Title: US-10-087-080-32

Perfect score: 2123

Sequence: 1 MKLEVFVRAHGHKQGSGL.....AALVRPGRPLSVPTLLA 402

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool.p/US10087080/runat\_29032004\_114947\_24501/app\_query.fasta\_1.583  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087080@cgn\_1\_103@runat\_29032004\_114947\_24501 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

- 1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	22.1	320	4	US-09-833-381-1266
2	468.5	22.1	3946	3	US-09-083-351-1
3	468.5	22.1	3946	3	US-09-083-352-1
4	446.5	21.0	1965	4	US-09-220-132-21
5	425	20.0	1659	3	US-09-083-351-3
6	425	20.0	1659	3	US-09-083-352-3
7	404.5	19.1	1860	2	US-08-331-644-3
8	404.5	19.1	1860	5	PCT-US93-04102-3
9	403.5	19.0	1309	4	US-09-576-594-927
10	400.5	18.9	2830	1	US-07-882-292-1
11	400.5	18.9	2830	2	US-08-331-644-1
12	400.5	18.9	2830	5	PCT-US93-04102-1

13	360.5	17.0	1155	2	US-08-331-644-4	Sequence 4, Appli
14	360.5	17.0	1155	5	PCT-US93-04102-4	Sequence 4, Appli
15	324.5	15.3	5080	4	US-09-976-594-495	Sequence 495, App
16	309.5	14.6	1605	4	US-09-087-134-10	Sequence 10, Appl
17	307	14.5	1634	4	US-09-087-134-13	Sequence 13, Appl
18	306.5	14.4	1793	3	US-09-113-309-1	Sequence 1, Appli
19	306.5	14.4	1793	3	US-09-521-109-1	Sequence 1, Appli
20	306.5	14.4	1793	4	US-09-562-332-1	Sequence 16, Appl
21	297	14.0	1668	4	US-09-087-134-16	Sequence 1, Appli
22	264	12.4	2517	1	US-07-906-930E-1	Sequence 1136, Ap
23	264	12.4	3465	1	US-09-023-655-1136	Sequence 3, Appli
24	263	12.4	2939	1	US-07-906-930E-3	Sequence 314, App
25	251.5	11.8	3394	4	US-09-620-312D-314	Sequence 3, Appli
26	236	11.1	3342	2	US-08-742-753-3	Sequence 1, Appli
27	233.5	11.0	3441	2	US-08-742-753-1	Sequence 1, Appli
28	231	10.9	441529	3	US-09-103-840A-1	Sequence 2, Appli
29	229.5	10.8	4403765	3	US-09-103-840A-2	Sequence 1, Appli
30	228.5	10.8	2888	4	US-08-765-907A-1	Sequence 1, Appli
31	226	10.6	441529	3	US-09-103-840A-1	Sequence 2, Appli
32	224	10.6	4403765	3	US-09-103-840A-2	Sequence 3, Appli
33	216	10.2	1869	4	US-09-372-668-3	Sequence 3, Appli
34	216	10.2	1869	4	US-09-697-377-3	Sequence 3, Appli
35	216	10.2	1869	4	US-09-696-868-3	Sequence 3, Appli
36	216	10.2	1869	4	US-09-697-341-3	Sequence 3, Appli
37	211.5	10.0	12001	1	US-08-458-568A-11	Sequence 35, Appl
38	210.5	9.9	2481	4	US-09-894-998A-35	Sequence 50, Appl
39	210.5	9.9	12425	4	US-09-616-289-50	Sequence 1, Appli
40	208	9.8	8438	1	US-07-945-283-1	Sequence 1, Appli
41	207.5	9.8	2150	2	US-08-318-837-1	Sequence 48, Appl
42	206.5	9.7	2561	4	US-09-616-289-48	Sequence 153, App
43	205.5	9.7	2109	4	US-09-370-838-153	Sequence 1, Appli
44	204	9.6	34094	4	US-09-292-034-1	Sequence 11, Appl
45	203	9.6	12001	1	US-08-458-568A-11	

ALIGNMENTS

RESULT 1

US-09-833-381-1266  
; Sequence 1266, Application US/09833381  
; Patent No. 6672186  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs  
; FILE REFERENCE: 5800-119  
; CURRENT APPLICATION NUMBER: US/09/833,381  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 09/516,448  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 2050  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1266  
; LENGTH: 320  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(320)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-833-381-1266

Alignment Scores:	1,09e-17	Length:	320
Pred. No.:	470.00	Matches:	94
Score:	89.62%	Conservative:	1
Percent Similarity:	88.68%	Mismatches:	9
Best Local Similarity:	22.14%	Indels:	2
Query Match:	4	Gaps:	0
DB:			

US-10-087-080-32 (1-402) x US-09-833-381-1266 (1-320)

Qy 190 AspsanTyrTrpMetLeuAnProAsnSerGluTyrThrPheAlaaspGlyValPheArg 209

Db 2 GACAACTACTGATGCTCAACCCCAAGAGAGTACCTTCGCCAGCGGGTCTTCGC 61  
Qy 210 ArgArgLysArgLeuSerHisArgAlaProValProAlaProGlyLeuArgProGlu 229  
Db 62 CGCGCGCGAAGCGCTCAACACCCGCGCGGTCCCGCGCGCGGTGGCGCGCGAG 121  
Qy 230 GluAlaProGlyLeuProAlaProProAlaProAlaProAlaProAlaSerProArg 249  
Db 122 GAGGCCCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181  
Qy 249 gMetArgSerProAlaArgGlnGluArgAlaSerProAlaGlyLysPheSerSerSe 269  
Db 182 CATGCGCTCG 241  
Qy 269 rPheAla-11eAspSerLleLeuArgLysProPheArgSerArgLeuArgAspThra 289  
Db 242 CTTTCATCATCTACAGATCTTCGCAATCTTCGCGATCGCGCGCGCGCGCGCG 301  
Qy 289 laProGlyThrThr 293  
Db 302 CCGCGCGGACGACC 315

## RESULT 2

US-09-083-351-1  
; Sequence 1, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,351  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UIA-029.02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3946 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 475..2133  
US-09-083-351-1  
Alignment Scores: 1.87e-16 Length: 3946  
Pred. No.:

Score: 468.50 Matches: 165  
Percent Similarity: 38.92% Conservative: 44  
Best Local Similarity: 30.73% Mismatches: 143  
Query Match: 22.07% Indels: 186  
DB: 3 Gaps: 23  
US-10-087-080-32 (1-402) x US-09-083-351-1 (1-3946)  
Qy 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGlnGlyAlaGlyGlySerAsp 27  
Db 287 CCGAGGCGCGGAGAGCCAGCCAGCGAGCGCGCGGAGAGCGGCGAGCGGCGAGCCGCG 346  
Qy 28 Ala-----ProSerProLeuSerAlaAlaGlyAspAspSerLeuGlySerAsp 43  
Db 347 GCACAGCGCAGCGCGCGCGGACCA---GCTCGCGCGCGCGCGGACTCGGACTCGCGCGCC 403  
Qy 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGlnGlyAspGly 63  
Db 404 GCGCGCGCGCGCGCGCGCGCGAGCGAGGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 463  
Qy 64 GluGlnSerAlaGlyGlyProGlyAlaGluAlaLleProAlaAlaAla----- 81  
Db 464 GAGCGG-----GGGCCA---TGCAGGCGCGCTACTCGGTCTCAGCGCCCAACT 508  
Qy 82 -----AlaAlaAlaValValAlaGluGlyAlaGluAlaGly 93  
Db 509 CCCTGGGAGTGGTCCCTACTCCTCGCGCGCGAGCAGAGTACTACCGC-GCGGCGCGCGCG 567  
Qy 94 AlaAlaGlyProGly----- 98  
Db 568 GCGCGCGCGCGCGCTACACCGCATCGCGCGCGCGCGCGCTACCGG-----CCCTAC 675  
Qy 99 -----AlaGlyGlyAlaGlySerGlyGlyAlaArgSerLysProTyr 113  
Db 628 CACGCCGAGCAGTACCGCGCGCGCATGGCGCGCGCTACGGG-----CCCTAC 675  
Qy 114 ThrArgArgPro-----LysProProTyrSerTyrLleAlaLle 127  
Db 676 ACGCCGCGCGCGCAGCCCAAGACATGGTGAACCGCGCTATAGCTACATCGCGCTCATC 735  
Qy 128 AlaMetAlaLleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluLleAsnGluTyr 147  
Db 736 ACCATGGCCATCCAGAACGCCCGCGGACAGATACCTGAGCGGCACTACCACTTC 795  
Qy 148 LeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSerValArg 167  
Db 796 ATCATGGACGGCTTCCCTTCTACCGGACACAAGCAGCGGCTGGCAGAACACATCCGC 855  
Qy 168 HisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArgProTyr 187  
Db 856 CACAACCTCTCGCTCAACGAGTGTCTTCAAGTGCCTCGCGCGCGCGCGCGCGCGCGCG 912  
Qy 188 GlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyVal 207  
Db 913 GGCAAGGCGCAGTACTGGACGCTGGACCGCGACTCTCTACACATGTTCGAGAACGCGCAGC 972  
Qy 208 PheArgArgArgLysArgLeuSerHis----- 217  
Db 973 TTCTGCGCGCGCGCGCGCGCTTCAAGAGAGAGCGCGGTGAAGCAAGGAGGAGAG 1032  
Qy 218 -----ArgAlaProValProAlaProGly 225  
Db 1033 GACAGGCTGCACCTCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1089  
Qy 226 LeuArgProGlnGluAlaProGlyLeuProAlaAlaPro----- 238  
Db 1090 -----CCGAGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1143  
Qy 239 -----ProProAlaProAlaProAla 246  
Db 1144 CAGGACATCAAGACCGAGAGCGGTACGTGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1203  
Qy 247 Ser-----ProArgMetArgSerPro----- 253

```
Db 1204 GCGCGCTGGGAGCGGAGCGCGCGCGGTCGCCAAGATCGAGAGCCCGCAGCAGC 1263
QY 254 ---AlaArgGlnGluArgAlaSerProAlaGlyLysPhe----- 266
Db 1264 AGCAGCAGCTTCCAGCGGAGCAGCGCGCGGCGAGCTCCGCTCGGCGGCGGCTC 1323
QY 266 ----- 266
Db 1324 AGCTGGAGCTGGGATTCGCGCGCGCGCGCGCGCGCTCCGCGCGCGCGCGCAG 1383
QY 267 ---SerSerSerPheAlaIleAspSerIleLeu-ArgLys----- 278
Db 1384 CATAGCAGGGCTTCAGCGTGACACATCATGACGTGCTCGGCGGCTCGCGCAGAGC 1443
QY 279 -----ProPheArgSerArgLeuArgAspThrAlaProG1 291
Db 1444 GCGGCGCGGAGCTCAGCTCCGCGCTTCGGCGCTCGGCGCGCGCTCCGCGCGG 1503
QY 291 YThrThrLeuGlnTrpGlyAlaAlaProCyBProProLeuProAlaPheProAlaLeuLe 311
Db 1504 ATCGACACCGCTGGCGTCTGGCGCTACTCGCGCGCGCAGAGCTCCCTCTACAGCTCC 1563
QY 311 uProAlaAla-----ProCyBArgAlaLeuLeuProLeuCyAlaTyArgLysG1 329
Db 1564 CCTGCGAGCAGCTCCAGCGCGGCGAGCTCGGCGCGCGCGCGCGCGCGCGG 1623
QY 329 uProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAla-----ProPr 346
Db 1624 G-----CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1674
QY 346 oLeuLeuLeuAlaProLeu-----ProAlaAlaAlaPr 357
Db 1675 AGCTGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1734
QY 357 oAlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyCysPro----- 374
Db 1735 GCCTGCGCGCTGGACACACCGCG-----TGCGCGAGTACTCT 1770
QY 375 ---LeuArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgPro 389
Db 1771 CTGCTCGCGTCCACGAGCAGCA-----GCTCGTCTCTCC 1804
```

## RESULT 3

```
US-09-083-352-1
; Sequence 1, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin W.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patel, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083.352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
```

```
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 475..2133
US-09-083-352-1
Alignment Scores:
Pred. No.: 1,87e-16 Length: 3946
Score: 468.50 Matches: 165
Percent Similarity: 38.92% Conservative: 44
Best Local Similarity: 30.73% Mismatches: 143
Query Match: 22.07% Indels: 186
DB: Gaps: 23
US-10-087-080-32 (1-402) x US-09-083-352-1 (1-3946)
QY 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGlnGlyAlaGlyGlySerAsp 27
Db 287 CCAGAGCGCGGAGGAGCAGCCCGCAGAGCGCGGAGAGCGCGGAGCGCGGAGC 346
QY 28 Ala-----ProSerProLeuSerAlaAlaGlyAspAspSerLeuGlySerAsp 43
Db 347 GCACAGCGAGCGGCGCGCGCACCA---GCTCGCGCGCGCGCGCGCGCGCGCGCG 403
QY 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGlnGlyAspGly 63
Db 404 GCGCGCGCGCGCGCGCGCGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463
QY 64 GluGlnSerAlaGlyGlyProGlyAlaGluGluAlaIleProAlaAlaAla----- 81
Db 464 GAGCGG-----GGGCGCA---TGCAGGCGCGCTACTCCGCTCCAGCGCGCGCACT 508
QY 82 -----AlaAlaAlaValAlaAlaGluGlyAlaGluAlaGly 93
Db 509 CCTGGGAGTGTGCTTACCTCGCGCGCGAGCAGAGTACTACCGC-GCGGCGCGCGG 567
QY 94 AlaAlaGlyProGly-----AlaGlyAlaGlySerGlyGlyAlaArgSerLysProTyr 113
Db 568 GCGCGCGCGGCGGCTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
QY 99 -----AlaGlyAlaGlySerGlyGlyAlaArgSerLysProTyr 113
Db 628 CAGCGCGAGCAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675
QY 114 ThrArgArgPro-----LysProProTyrSerTyrIleAlaLeuLeu 127
Db 676 ACGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735
QY 128 AlaMetAlaIleArgAspSerAlaGlyArgLeuThrLeuAlaGluLeuAlaGluTyr 147
Db 736 ACCATGGCCATCCAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 795
QY 148 LeuMetGlyLysPhePheArgGlySerTyrThrGlyTyrArgHisSerValArg 167
Db 796 ATCATGGACCGCTTCCCTTCTACCGGAGACAAAGAGCGGCGCGCGCGCGCGCGCG 855
QY 168 HisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArgProTyr 187
Db 856 CACAACCTCTCGTCAACGAGTGTCTCGTCAAGGTCGCGCGCGCGCGCGCGCGCGCG 912
QY 188 GlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyVal 207
```

```
Db 913 GGAAGGCGAGCTACTGGAGCTGGACCGGAGCTCTACACATGTTGAGAACGCGCAGC 972
Qy 208 PheArgArgArgArgArgArgArgArgArgArgHis- 217
Db 973 TTCTGCGGCGCGCGCGCGCTTCAAGAAAGAGACGCGGTGAAGACAAGGAGGAGAAG 1032
Qy 218 -----ArgAlaProValProAlaProGly 225
Db 1033 GACAGGCTGACCTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1089
Qy 226 LeuArgProGluGluAlaProGlyLeuProAlaAlaPro----- 238
Db 1090 -----CCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATC 1143
Qy 239 -----ProProAlaProAlaAlaProAla 246
Db 1144 CAGGACATCAAGACCGAGAACGTAAGTGCCTCGCGCGCGCGCGCGCGCGCGCGCG 1203
Qy 247 Ser-----ProArgMetArgSerPro----- 253
Db 1204 GCGCGCTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1263
Qy 254 -----AlaArgGlnGluArgAlaSerProAlaGlyLysPhe----- 266
Db 1264 AGCAGCAGCTGTCCAGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1323
Qy 266 ----- 266
Db 1324 AGCTGACGGTGGGATTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1383
Qy 267 -----SerSerSerPheAlaIleAspSerIleLeu-ArgLys----- 278
Db 1384 CATAGCAGCGCTTACGTGGACAAACATCATGCTGCGCGCGCGCGCGCGCGCGCG 1443
Qy 279 -----ProPheA-ArgSerArgArgLeuArgAspThrAlaProGln 291
Db 1444 GCGCGCGCGAGCTACGCTCGCGCTTCTGGCTCGCGCGCGCGCGCGCGCGCGCG 1503
Qy 291 YThrThrLeuGlnTrpGlyAlaAlaProCysProProLeuProAlaPheProAlaLeu 311
Db 1504 ATCGACCGCGCTGGCGCTCGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCG 1563
Qy 311 UProAlaAla-----ProCysArgAlaLeuLeuProLeuProCysAlaTyrGlyAlaGln 329
Db 1564 CCCTGACGACGAGCCTCCAGCGCGGAGCTCGCGCGCGCGCGCGCGCGCGCGCG 1623
Qy 329 UProAlaArgLeuGlyAlaArgGlnAlaGluValProProThrAla-----ProPr 346
Db 1624 G-----CGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATG 1674
Qy 346 oLeuLeuAlaProLeu-----ProAlaAlaAlaPr 357
Db 1675 AGCTGTACGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1734
Qy 357 oAlaAlaProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysPro----- 374
Db 1735 GGCTCGCGCGTGGACAACCGCG-----TGCGCGCTACTCT 1770
Qy 375 ----LeuArgLeuProAlaAlaLeuGlnAlaLeuValArgPro 389
Db 1771 CTGCTCGCGTCAACAGCAGCA-----GCTCGTCTGCTCC 1804

RESULT 4
US-09-220-132-21
; Sequence 21, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE OF INVENTION: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
```

```
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-220-132-21

Alignment Scores:
Pred. No.: 1,31e-15 Length: 1965
Score: 446.50 Matches: 150
Percent Similarity: 44.55% Conservative: 42
Best Local Similarity: 34.80% Mismatches: 121
Query Match: 21.03% Indels: 120
DB: 4 Gaps: 19

US-10-087-080-32 (1-402) x US-09-220-132-21 (1-1965)
Qy 28 AlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAspGlyAspCysAla 47
Db 13 GCGCGGAGACACCTCTCTCTCC-----TCGTCGCTCTCTCGCTCTCTCGGCC 63
Qy 48 AlaLysProSerAlaGlyGlyGlyAlaArgAspThrGlnGlyAspGlyGluGlnSerAla 67
Db 64 TCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 90
Qy 68 GlyGlyGlyProGlyAlaGluGluAlaIleProAlaAlaAlaAlaAlaValAlaAla 87
Db 91 -----AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 114
Qy 88 GluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyGlyAlaGlySerGlyGluGly 107
Db 115 -----AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 156
Qy 108 AlaArgSerLysProTyrThrArgArgPro-----LysProProTyrSerTyrIleAlaLeu 126
Db 157 AAGAGGCGAGCTCGGGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 216
Qy 127 IleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGlu 146
Db 217 ATCGTCATGGCCTCAGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 276
Qy 147 TyrLeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSerVal 166
Db 277 TTCCTGCGCGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 336
Qy 167 ArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArgPro 186
Db 337 GCGCAATCTCTCGCTCAACGAGTCTTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 396
Qy 187 TrpGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGly 206
Db 397 ---GSCAAGGCGCACTACTGGACCATCGACCGCGCGCGCGCGCGCGCGCGCG 453
Qy 207 ValPheArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 221
Db 454 TCGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 512
Qy 222 Pro-----AlaProGlyLeuArgPro 228
Db 513 CCACCGCGTGTGAGCGGCTTGGGCTTCGGGCGCTGCTGCTGCTGCTGCTGCTGCT 569
Qy 229 GluGluAlaProGlyLeuProAlaAlaAlaProProAlaProAlaAlaProAlaSerPro 248
Db 570 -----CTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 614
Qy 249 ArgMetArgSerPro-----AlaArgGlnGlu 257
```

Db 615 CGGCTACGGCGCTCGACATGATGCCCGGGGTACGACGCCGGCGGGCGGGCCCCCAG 674  
Qy 258 GluArgala-----SerProAlaGlyLysPheSerSerSer 269  
Db 675 CCACGCGACCTCCACACACCACCACCACCGTCCGCACATGTCGCCCAACCCGGG 734  
Qy 270 PheAlaIleAspSerIleLeu-----ArgLysProPheArg 281  
Db 735 TTCACCTACATGGCAGCTGCCGGTGGCCCGGGAGCCCGGGGCGTGGTGGCGCGG 794  
Qy 282 SerArgLeuArg-AspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaProCy 301  
Db 795 GGGCGGCGCGCGCGCTACGACGCGCGCAGCAGCAGC-----ACGACGCCGT 842  
Qy 301 sProProLeuProAlaPhe-----ProAlaLeuLeuProAl 313  
Db 843 ACCCTCGTCCCGGCGCATGGGAGCGCCATGAATGCCATCGCCCTACACGAGCCCTGC 902  
Qy 313 aAla-----ProCysArgAlaLeuLeuPro-----LeuCysAlaTyrG 326  
Db 903 GCGGCACTGGAGCTCGCTCGG-CGCTCGCTTACTCAAGCAGCGCGCTCGCTGAGCG 961  
Qy 326 yAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAlaProPr 346  
Db 962 CCAGCAGCAACCCCGCGCTCGGCGAG-----CCTGCACCTCCAGCATGT 1006  
Qy 346 oLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAl 366  
Db 1007 CCTCTACTCGCTGAGCAGAGCTACTTGACACGAGCTCG---CGAGGACTCTCAG 1063  
Qy 366 adGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaLeu-- 385  
Db 1064 TGGGACTGCC-----CCGTTACAGCATCACTTACTTCCAGTGTGTG 1105  
Qy 386 -ValArgArgProGlyProHisLeuSerTyr 395  
Db 1106 ACAGAAAGATTTCGCTCTCAATTCATGG 1136

RESULT 5  
US-09-083-351-3  
; Sequence 3, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patel, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083.351  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Atgold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UTA-029.02  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1659 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-09-083-351-3  
Alignment Scores:  
Pred. No.: 1-51e-14 Length: 1659  
Score: 425.00 Matches: 140  
Percent Similarity: 39.51% Conservative: 40  
Best Local Similarity: 31.04% Mismatches: 112  
Query Match: 20.02% Indels: 159  
DB: Gaps: 19  
US-10-087-080-32 (1-402) x US-09-083-351-3 (1-1659)  
Qy 72 GlyAlaGluAlaIleProAlaAlaAlaAlaAlaValAlaGluGlyAlaGlu 91  
Db 58 GGGCGGCGAGCAGAGCTACTACCGCGCGCGCGCGCGCG-----GCCGGGGCGCGCTAC 111  
Qy 92 AlaGlyAlaAlaGlyProGlyAla----- 99  
Db 112 ACGCCATCGCGCGCCCATGAGCGTGTACTCGCACCTTGGCACCGCGCGAGTACCGG 171  
Qy 100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgPro----- 117  
Db 172 GGGCGCATGCGCGCGCTACGCG-----CCCTACACGCGCGCGCGAGCCCC 219  
Qy 118 -----LysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAsp 133  
Db 220 AAGGACATGGTGAAGCGCGCTTACTATCATCGCGCTCATCACCATGCGCATCCAGAAC 279  
Qy 134 SerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPhePro 153  
Db 280 GCGCGGACAAAGATGATCCCTGACGCGCATCTACAGTTTCATGAGCGCTTCCCG 339  
Qy 154 PhePheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsn 173  
Db 340 TTCTACCGGACAAAGCAGCGGCTGGCAGACAGATCCGCCAACACCTCTCGTCTCAAC 399  
Qy 174 AspCysPheValLysValLeuArgAspProSerArgProTyrGlyLysAspSerTyrTrp 193  
Db 400 GAGTGTCTCGTCAAGTGGCGCGAGCAGACAAAGACCG-----GGCAAGGCGAGCTACTGG 456  
Qy 194 MetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLys 213  
Db 457 ACCTGGACCGGAGCTCTTACAAACATGTTTCGAGAACGCGAGCTTCTCTGGCGCGCGG 516  
Qy 214 ArgLeuSerHis----- 217  
Db 517 CGCTTCAAGACGAGGACGCGGTGAAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 576  
Qy 218 -----ArgAlaProValProAlaProGlyLeuArgProGluGluAla 231  
Db 577 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG-----CCGAGGAGCGCC 627  
Qy 232 ProGlyLeuProAlaAlaPro----- 238  
Db 628 GACGGCAACG 687  
Qy 239 -----ProProAlaProAlaProAlaSer----- 247  
Db 688 AACGGTACGTGCG 747  
Qy 248 -----ProArgMetArgSerPro-----AlaArgGlnGluGlu 258  
Db 748 AGCGCGCGCGCGGTGCCAAGATCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807

```
QY 259 ArgAlaSerProAlaGlyLysPhe-----266
Db 808 GGGAGCAGACCCCGGGCAGCGCTCGCGCGCGCGCGCTCAGCCTGGAGCGGTGCGAT 867
QY 267 -----SerSerSerPheAla 271
Db 868 TCCGCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 927
QY 272 lleAepSerileLeu-ArgLys-----278
Db 928 GTGGACAACATCATGATCGCTCGCGCGCGGTGCGCGAGCGCGCGCGCGAGCTCAGC 987
QY 279 ----ProPheArgSerArgAlaGlyLeuArgPheAlaProGlyThrThrLeuGlnTrpGI 297
Db 988 TCCGCGCTTCGGGCTCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1047
QY 297 yAlaAlaProCysProProLeuProAlaPheProAlaLeuProAlaAla-----Pr 315
Db 1048 CTCGGCGCTACTCGCGCGCGCGAGCTCCCTCTACAGCTCCCGCTGACGACGACCTCC 1107
QY 315 cCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAl 335
Db 1108 AGCGCGGCGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1158
QY 335 aArgGluAlaGluValProProThra-----ProProLeuLeuAlaProLe 352
Db 1159 GCGGCGCGCGCGCGGAGCTTACCACTGCAACTGCAAGCCATGAGCTGTACGCGCGCGCG 1218
QY 352 u-----ProAlaAlaProAlaLysProLeuArgGI 363
Db 1219 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1278
QY 363 yProAlaAlaGlyAlaHisLeuTyrCysPro-----LeuArgLeuProAl 379
Db 1279 CCCC-----TGCCGAGTACTCTCTGCTCGCTCCGTCACGAGC 1314
QY 379 aAlaLeuGlnAlaAlaLeuValArgArgPro 389
Db 1315 AGCA-----GCTGCTGCTGCC 1330

RESULT 6
US-09-083-352-3
; Sequence 3, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patti, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, ROAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1659 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-083-352-3

Alignment Scores:
Pred. No.: 151e-14 Length: 1659
Score: 425.00 Matches: 140
Percent Similarity: 39.91% Conservative: 40
Best Local Similarity: 31.04% Mismatches: 112
Query Match: 20.02% Indels: 159
DB: 3 Gaps: 19

US-10-087-080-32 (1-402) x US-09-083-352-3 (1-1659)
QY 72 GlyAlaGluGluAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlu 91
Db 58 GCGCGCGAGCAGAGCTACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 111
QY 92 AlaGlyAlaAlaGlyProGlyAla-----99
Db 112 ACCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 171
QY 100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgPro-----117
Db 172 GCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 219
QY 118 -----LysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAsp 133
Db 220 AAGGACATGGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279
QY 134 SerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPhePro 153
Db 280 GCGCGCGAGAGAGATACCGCTGAAACGGCATCTACCGATTCATCGGACCGCTTCCCC 339
QY 154 PhePheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsn 173
Db 340 TTCTACCGGACAAACAGCAGCGGTGGCAGAACAGCATCCGCCACAACTCTCGCTCAAC 399
QY 174 AspCysPheValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrp 193
Db 400 GAGTCTTCGTCAAGGTCCCGCGCGAGCAAGAGCGCG---GGCAAGGCGCGCTACTGG 456
QY 194 MetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLys 213
Db 457 ACGCTGGACCGGACTCCTACACATGTTTCGAGAACCGGAGCTTCCTCGCGCGCGCGG 516
QY 214 ArgLeuSerHis-----217
Db 517 CGCTTCAAGAAAGAGGACGCGGTGAAGCAAGAGGAGAGGAGGAGGAGGAGGAGGAGG 576
QY 218 -----ArgAlaProValProAlaProGlyLeuArgProGluGluAla 231
Db 577 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
QY 232 ProGlyLeuProAlaAlaPro-----238
Db 628 GACGGCAACGCGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687
QY 239 -----ProAlaProAlaAlaProAlaSer-----247
Db 688 AACGGTACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
QY 248 -----ProArgMetArgSerPro-----AlaArgGlnGlu 258
Db 748 AGCGCGCGCGCGGTGCGCGCAAGATCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
```



Db 953 CAACCGG---GGCAAGGCGCACTACTGGAGCGCTCGACCCCGGAGTCCCGACATATGTTGGA 1009  
Qy 204 aAspGlyValPheArgArgArgLysArgLysHisArgAlaProValProAlaPr 224  
Db 1010 CAACGGAGCTTCTCGGGCGCCGCAAGCGTTCAAG---CGCAGCGCTACTCGATCC 1066  
Qy 224 oGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaAla 244  
Db 1067 CGCGCTG----- 1073  
Qy 244 aProAlaSerProArgMetArgSerProAlaArgInGluGluArgAlaSerProAlaGl 264  
Db 1074 -----GGGACTCTGCACCAAGGACAGCGCTGCCAGTGTGA 1111  
Qy 264 yLysPheSerSer-SerPheAlaIleAspSerIleLeuArgLysProPheArgSerArgA 284  
Db 1112 GAATTTACTCTAGGATTT-----CCAAATTGAGAACGCTGTAGCGCGGG 1162  
Qy 284 rgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGly-----AlaAlaProCysP 302  
Db 1163 AGAGCGAAGGTAGGACTCCCGCTTCTTCTCCGATGGGGGTTGGTTTCTGTCGCC 1222  
Qy 302 roProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuPro- 321  
Db 1223 CTCCCGGTCTTCGGAGACCCCGCGCCCG---CGTTCCTCCCGCTTCGGATTCTGGACCAG 1281  
Qy 322 -LeuCysAlaTyrgLysAlaGlyGluProAlaArgLeuGlyAlaArg 336  
Db 1282 ACTGTGTGGCGCAGAGCTGGGGCGCGCGAGTTAGTCTCAGAGG 1327

## RESULT 8

PCT-US93-04102-3  
Sequence 3, Application PC/TUS9304102  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04102  
FILING DATE: 19930430  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,292  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41472A-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-977-9550  
TELEFAX: 212-664-0525  
TELEX: 423523 COOP UI  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1860 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA

## PCT-US93-04102-3

Alignment Scores:  
Pred. No.: 2,09e-13 Length: 1860  
Score: 404.50 Matches: 133  
Percent Similarity: 40.95% Conservative: 30  
Best Local Similarity: 33.42% Mismatches: 132  
Query Match: 19.05% Indels: 104  
DB: 5 Gaps: 13  
US-10-087-080-32 (1-402) x PCT-US93-04102-3 (1-1860)  
Qy 8 ProArgAlaAlaHis-----GlyAspLysGlnGlySerAspLeuGluGly 22  
Db 245 CGCGCGCTCGCAGAGTGGCTCTCGCTCCGGTCCGCCCTCCGGATCGCCTCGGGA 304  
Qy 23 AlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAsp----- 37  
Db 305 GGCAGGGAGCGCGGAGCGCGTCCCTATGTGCCCGCGCCACCGCGCGGCCCCAGC 364  
Qy 38 ---AspSerLeuGlySerAspGlyAspCys----- 46  
Db 365 TATGACCTGACGACGAGATGTCGATGCTCCGGCTCGGGAGGAGACAGACATCGA 424  
Qy 47 -----AlaAlaLysProSerAlaGlyGlyAlaArg-AspThrGlnGlyAspGl 63  
Db 425 CGTGTGGGGAGCGGAGGACGACGAGGAGGAGGACGATGACGAGGGCGCGG 484  
Qy 63 yGluGlnSerAlaGlyGlyGly----- 70  
Db 485 CGCGCGCGCGCGCGCGGTCCCGTTCGCGAGTCCGCCCGCGCGGCGGCGCTCTTA 544  
Qy 70 ----- 70  
Db 545 CGCGCGGAGGTGCGATCTCGAGGAGTGGAGGAGGAGGACGATGATGCTGCTCGC 604  
Qy 71 -----ProGlyAlaGluAlaIleProAlaAlaAlaAlaAlaValAlaGl 88  
Db 605 CCGCGCGCGCGCGCTCCCGCGCTCCGGTCTCGCGCGCGCGCGCGGAGTGTAGCGG 664  
Qy 88 uGlyAlaGluAlaGlyAla---AlaGlyProGlyAlaGly-----GlyAlaGlySe 104  
Db 665 GGGCGCTCGACGCGCGCGGAGCGGAGCGCGGCGCGGAGTGTAGCGGCGCGGCGAC 724  
Qy 104 rGlyGluGlyAlaArgSerLysProTyThrArgArgProLysProTyThrSerTyri 124  
Db 725 GGGCGGGCGCTAAGAT---CCGCTGGTG-----AAGCGCGCTACTCGTACAT 772  
Qy 124 eAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGlu 144  
Db 773 CGCGCTCATCACATGGCCATCTCTCAGAGCGCCCAAGAGCGCTGACGCTCAGCGAG 832  
Qy 144 eAsnGluTyLeuMetGlyLysPheProPheArgGlySerTyThrGlyTyThrArgAs 164  
Db 833 CTGCGATTTCATCAGCAGCGCTTCCCTTACTACCGGAGAGTTCGCCGCTGGCAGAA 892  
Qy 164 nSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLysArgAspProSe 184  
Db 893 CAGCATCCGTCAACCTGTGCTCAACAGCTGCTTCTCAAGATCCCGCGCAACCGGG 952  
Qy 184 rArgProTrpGlyLysAspAsnTyTrpMetLeuAsnProAsnSerGluTyThrPheAl 204  
Db 953 CAACCCG---GGCAAGGGCAACTACTGACCGCTCCCGGAGCGTTCAG--CGCCAGCGCTACTCGATCC 1009  
Qy 204 aAspGlyValPheArgArgArgLysArgLysHisArgAlaProValProAlaPr 224  
Db 1010 CAACGGAGCTTCTCGGGCGCCGCAAGCGTTCAAG---CGCCAGCGCTACTCGATCC 1066  
Qy 224 oGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaAla 244  
Db 1067 CGCGCTG----- 1073  
Qy 244 aProAlaSerProArgMetArgSerProAlaArgInGluGluArgAlaSerProAlaGl 264

Db 1074 -----GGACTCTGCACCAAGGACAGCGCTGTCAGTGTGGA 1111  
 QY 264 YLysPheSerSer-SerPheAlaIleAspSerIleLeuArgLysProPheArgSerArgA 284  
 Db 1112 GAACCTTTACTGCTAGGATTT-----CCAAATTGTTAGGAACGCTGTAGCGCGGG 1162  
 QY 284 rGLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGly-----AlaAlaProCysP 302  
 Db 1163 AGAGCGAGGTAGACTCCCGGCTCTTCTCCGATGGGGGGTGGTTTGGTTCGCCCC 1222  
 QY 302 roProLeuProAlaPheProAlaLeuLeuProAlaProCysArgAlaLeuLeuPro- 321  
 Db 1223 CTCCTCGGCTCTCGAGACCCCGCGGCC-CCGTTTTCGGCGCTTCGGATTCTTGACACAG 1281  
 QY 322 -LeuCysAlaTyrglyAlaGlyGluProAlaArgLeuGlyAlaArg 336  
 Db 1282 ACTGTGTGGCGCACAGCTGGGCGCGCGCAGCTTTAGTCTCAGG 1327

## RESULT 9

US-09-976-594-927  
 ; Sequence 927, Application US/09976594  
 ; Patent No. 6673549  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Furness, Michael  
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
 ; FILE REFERENCE: PA-0041 US  
 ; CURRENT APPLICATION NUMBER: US/09/976,594  
 ; CURRENT FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: 60/240,409  
 ; PRIOR FILING DATE: 2000-10-12  
 ; NUMBER OF SEQ ID NOS: 1143  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 927  
 ; LENGTH: 1309  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; OTHER INFORMATION: Incyte ID No. 6673549 3687719CB1  
 US-09-976-594-927

## Alignment Scores:

Pred. No.: 1,63e-13 Length: 1309  
 Score: 403.50 Matches: 135  
 Percent Similarity: 44.21% Conservative: 33  
 Best Local Similarity: 35.53% Mismatches: 136  
 Query Match: 19.01% Indels: 77  
 DB: 4 Gaps: 14

US-10-087-080-32 (1-402) x US-09-976-594-927 (1-1309)

QY 75 GluAlaIleProAlaAlaAlaAlaAlaValAlaAla-GluGlyAlaGluAlaGlyAl 94  
 Db 4 GAAGCTGAGCCTGACCCAGCAGGTCCAGCAGCCTGGCAGCCCGCCAGCATGCAGCAGC 63  
 QY 94 alaGly-ProGlyAlaGlyGlyAlaGlySerGlyGlyAlaArgSerLysProTyrT 114  
 Db 64 AGCCTCTGCCCGGCTGCGCC-----CCCAAA 93  
 QY 114 hrArgArgProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAsps 134  
 Db 94 CTGAGCCACCAAGCCTCCCTACAGCTACATCGCCCTATTGCTATGCCATCCAGACT 153  
 QY 134 eAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProp 154  
 Db 154 CACCGGGGAGGGGCGCCACCTCTCAGTGCATCTACCGGTACATCGGCGGCAATTCGCT 213  
 QY 154 hPheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnA 174  
 Db 214 TCTACGGCACACCGGCGCGGCTGGCAGACAGCATCCGCCAATCTGCTCACTCAACG 273

QY 174 sPCysPheValIysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpM 194  
 Db 274 AGTGTCTTTGTCAAGGTGCCCGCATGACCGCAAGCCA---GGCAAGGCGGCTACTGGA 330  
 QY 194 eLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgArgLysA 214  
 Db 331 CGCTGACCTTGACTGCCACACACATGTTTGACGACGGCAGCTTCTACGCGCGCGCCGCC 390  
 QY 214 rGLeuSerHis-ArgAlaProAlaProAlaProGlyLeuArgProGluGluAla----- 231  
 Db 391 GCTTCACCGCGCAGACAGGTGCTGAGGACACCGCGGCGCCGCAAGGACGCGCGTGAC 450  
 QY 232 ---ProGlyLeuProAla-----AlaProProAlaProAlaProAlaProAla 246  
 Db 451 CCTCAGGCGCAGCAGCAGCAGCAGGAGTCCCAAGCGCACCGCAGCGGCGGAGTGCT 510  
 QY 247 SerProArgMetArgSerProAlaArgGlnGluAlaArgAla----- 260  
 Db 511 CATTCGCCACACAG-----AGCTGCCAGATCCCAAGGGCCTAAGCTTTGGGGGTCTGG 561  
 QY 261 ---SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysPro 279  
 Db 562 TGGGGGCGCATCCAGCCAGTATGTGCCAGCAACCACTGATGGCAGGCTCGGCCACCCA 621  
 QY 280 PheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGln---TrpGlyAla 298  
 Db 622 TGGAGCCCAAGAGATTT-----CCAGCCCAAGCCTGCATGCCAGGGGAGC 669  
 QY 299 AlaProCysProPro---LeuProAlaPheProAlaLeu---LeuProAla----- 313  
 Db 670 TCCCGGTGGCCACCTCATCTCTCTCA-TGCCACGCGTTTGGCTTTCTGCGCGGCTTCTCA 728  
 QY 314 -----AlaProCysArgAlaLeuLeuProLeuProLeuCysAlaTyrgly 326  
 Db 729 GAGCTGAGAGTTTATTAAGGCCCTACCGCCGCTTGTCTCCCGGATCAGGCATCGG 788  
 QY 327 AlaGlyGluProAlaArgLeu----- 333  
 Db 789 AGCAGTACCAGTGTCTGGCTGCAGGCACTGAATTTTGCATGGGGGTGACCCAGGCGCT 848  
 QY 334 -----GlyAlaArgGluAlaGluValProProThrAlaProProLeuLeu 348  
 Db 849 GAGCAGCTTTGGCCTCAGCAGCCCTCCCTGACCCACCCCTCCAGGCTCACTC 908  
 QY 349 LeuAlaProLeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAlaGlyGly 368  
 Db 909 CGGCGCCCACTGCGC-----CTGCCCACTGACCAAGGACCCCTGGGTGAGGTGGC 962  
 QY 369 AlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaAlaLeuValArg 388  
 Db 963 TTCCTGTCCAGGAGGCTCCGGCTACCCATTGGGGCTGACCCCTGCTATACCGGAGC 1022  
 QY 389 ProGly-----ProHisLeuSerTyrProValGlu 398  
 Db 1023 CCAGGAATGTTCTTCTTGTAGTAAGACGACCTCACTCGGGCAGTCCCTGCGAG 1076

## RESULT 10

US-07-982-292-1  
 ; Sequence 1, Application US/07882292  
 ; Patent No. 5324638  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tao, Mufan  
 ; APPLICANT: Lai, Eseng  
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS  
 ; TITLE OF INVENTION: ENCODING SAME AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White  
 ; STREET: c/o Cooper and Dunham, 30 Rockefeller  
 ; STREET: Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA





QY 375 uArgLeuProAlaLeu-----GlnAlaLeuValArgArgProGlyPr 391  
 DB 1674 ACCT---CCGACCGTCAATGACTTCGACAGCAGCAGCTCCATGAGCGCCCGCGCG 1730  
 QY 391 cHisLeuSerTyr 395  
 DB 1731 CGTCTCTCTCTAC 1743

## RESULT 12

PCT-US93-04102-1  
 ; Sequence 1. Application PC/TUS9304102  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tao, Wufan  
 ; APPLICANT: Lai, Beeng  
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White  
 ; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/04102  
 ; FILING DATE: 19930430  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/882,292  
 ; FILING DATE: 13-MAY-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 41472A-PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-977-9550  
 ; TELEFAX: 212-664-0525  
 ; TELEX: 422523 COOP UI  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2830 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; HYPOTHEetical: N  
 ; ANTI-SENSE: N  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 443..1882  
 ; OTHER INFORMATION:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 926..1255  
 ; OTHER INFORMATION: /note= "nucleotide sequence encoding DNA  
 ; OTHER INFORMATION: binding domain homology"  
 ; FEATURE:  
 ; NAME/KEY: misc\_signal  
 ; LOCATION: 1883..1885  
 ; OTHER INFORMATION: /note= "translation termination codon"  
 PCT-US93-04102-1

Alignment Scores:  
 Pred. No.: 5,32e-13 Length: 2830  
 Score: 400.50 Matches: 143  
 Percent Similarity: 40.47% Conservative: 29  
 Best Local Similarity: 33.65% Mismatches: 146

Query Match: 18.86% Indels: 109  
 DB: 5 Gaps: 18  
 US-10-087-080-32 (1-402) x PCT-US93-04102-1 (1-2830)  
 QY 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27  
 DB 683 CCCCCGACGGCGCGCGCGCCCGCCAGCAGGGGACGACGACGAGGGC----- 727  
 QY 28 AlaProSerProLeu-----SerAlaAlaGlyAsp-----AspSer 39  
 DB 728 ---CCCCAGCCGCTTCTGCTCCCGCCGCTCCCGCCGCTGAGCGGCGCAAGCTGACGCA 784  
 QY 40 LeuGlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThr 59  
 DB 785 CTTGGAGCCAAAGGCGAG-----CCAGCGCGCGCGCGCTCGGAGCTGGCGCC 832  
 QY 60 GlnGlyAspGlyGluLysSerAlaGlyGlyGlyProGlyAlaGluGluAlaAlaProAla 79  
 DB 833 GTCGGCGCGGACGAGAGGAGGCGCGCGCGCTGGGGGGGAGGAG----- 880  
 QY 80 AlaAlaAlaAlaAlaAlaAlaAlaGluGlyAlaAlaAlaGlyProGlyAla 99  
 DB 881 -----AAGAAGGGCGCGCGCGAGGGCGGCAAG 907  
 QY 100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysPro 119  
 DB 908 GACGGGGAGGGGGCGCAAGAGGGC-----GACAGAGAACACGCGCAAGTACGAGAACCG 961  
 QY 120 ProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeu 139  
 DB 962 CGTTTCACTCAACCGCTCATCATGATGCTCATGAGGAGGAGGAGGAGGAGGAGGAGGAG 1021  
 QY 140 ThrLeuAlaGluIleAsnGluTyrLeuMetClyLysPheProPheArgGlySerTyr 159  
 DB 1022 AGCTCAAGCGCATCTACGATTTTCATGAGAACTTCTTACTACCGCGAGAACAG 1081  
 QY 160 ThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValVal 179  
 DB 1082 CAGGGCTGGGAGAACTCCATCCGCAACACCTGTCTCCCTCAACAGTCTTCGTGAGGTA 1141  
 QY 180 LeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSer 199  
 DB 1142 CGCGCCCACTAGCAGGACCCG---GGCAAGGCGCACTACTGATCTGGACCGCTCGAGC 1198  
 QY 200 GluTyrThrPheAlaAspGlyVal-----PheArgArgArgLysArgLeuSer 216  
 DB 1199 GACGACGTGTTTCATCGCGCGGCGACCGCGCAAGCTGCGGCGCGCTC-CACCACTCTCG 1257  
 QY 217 HisArgAlaProValProAlaProGlyLeuArgProGluAlaProGlyLeuProAla 236  
 DB 1258 GGCCAGCTAGCTTTAAGCGCGCGG-----ACG 1287  
 QY 237 AlaProProAlaProAlaProAlaProAlaProArgMetArgSerProAlaArgGln 256  
 DB 1288 GCTCA-CCTCCA-----CCGCGCTTCACTTATGAGCGCGCGCGCT----- 1328  
 QY 257 GluGluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeu 276  
 DB 1329 ---CCCTCTACTGCGCCATGT----- 1346  
 QY 277 ArgLysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrp 296  
 DB 1347 -----CGCCCTTCTGTCTCCCGACCACTCCCTCGCGGCA 1379  
 QY 297 GlyAla-----AlaProCysProProLeuProAlaPheProAlaLeuLeu 311  
 DB 1380 CGAGCACTTTCAGTTTCAACGGGACCACTCGCGCTTACCCAGCCACCCCATGC----- 1433  
 QY 312 ProAlaAlaProCys-----ArgAlaLeuLeuProLeu----- 322  
 DB 1434 CCTACAGCTCGTGTGACTCAAAACTCGCTGGGGCAACCACTCTCTCTCCACCGGCA 1493



; GENERAL INFORMATION:  
 ; APPLICANT: Tao, Wufan  
 ; APPLICANT: Lai, Eseng  
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:

Alignment Scores:		
Pred. NO.:	2.72e-11	Length:
Score:	360.50	Matches:
Percent Similarity:	45.27%	Conservative:
Best Local Similarity:	34.32%	Mismatches:
Query Match:	16.98%	Indels:
DB:	5	Gaps:
		13
		1.155
		1.6
		37
		1.8
		67
		13

US-10-087-080-32 (1-402) x PCT-US93-04102-4 (1-1155)

QY	26	SerAspAlaProSerPro---LeuSerAlaAlaGlyAspSerLeuGlySerAspGly	44
Db	207	TCCTGCTCTCTCGGGCTCGCTCGCGCGCGCAGTCGTCGCGCAGCGCGC	266
QY	45	AspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAspGlyGlu	64
Db	267	GACCGGCGCTCTGGGACACAGCTGCGCGCG-----TTACTCGGTATCGACCC	317
QY	65	GlnSerAlaGlyGlyGly-ProGlyAlaGluGluAlaIleProAlaAlaAlaAlaAla	84
Db	318	CAACACCTCGGAGTGGTACCTATTTGAGTGACAAACTACTACGGCGCGCGCGAG	377
QY	84	avalValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyAla	102
Db	378	CTAC-----GGCGGATGCCAGCCCATGGGGTCTACTCCGCCACCCGGAGCA	428
QY	103	----GlySerGlyGluGlyAlaArgSerLysProTyrThrArgPro-----	117
Db	429	GTACGGCGCGCATGGGCGCTCTACGCGCCCTTACCACCATCAGCCCTTTTCTCCCA	488
QY	118	----LysProTrpTyrSerTyrIleAlaIleAlaIleAlaIleArgAspSe	134

489	GGACCTGGTGAAGCCGCCCTCATCAGCTATATAGCGTCTATCACCATGGCGCATCCAGAACCG	548
134	rAlaGlyGlyArgLeuThrLeuAlaGluLeuAsnGluTyrLeuMetGlyLysPheProPh	154
549	GCACAGAAAGAGATCATTCTGAAAGGGCATACACAGTTCATCATGTGACCGTTTCCCTT	608
154	ePheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAs	174
609	CTACCGCGAGAACACAGCAGGGCTGGCAGAACAGCATCCGCCACAACCTGTCTACTCAATGA	668
174	pCysPheValIysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMe	194
669	GTGCTTCGTGAAGTCCCGCGCAGACACAGAACCGCG--GCACAGGCGCATCTCTGGAC	725
194	tLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArg	214
726	GCTCGACCCGGACTCCTACAACATGTTGAGAATGGCAGTCTCTCTCGCGCGCGCGGCG	785
214	gLeuSerHisArgAlaProValProAlaPArgLysLeuArgProGluGlu-AlaProGlyL	234
786	CTTCAAG---AAGAAGGATGTGCCCAAGAC-----AAGGAGGAGCGGCCCCACCT	833
234	euProAlaAlaProProAlaProAlaAla-----	244
834	CAAGGAGCGCCCTCGACCACCGCCAAAGGCGCTCCGACAGGACCCCGTAGTCAACGG	893
245	--ProAlaSerProArgMetArgSerProAlaArgGlnGlu-----A	259
894	GCCCAAGGAGGCGCAGAGAAAGATCGTGTTAAGACGAGGCGGCGTCCCGCGCATGCC	953
259	rgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysP	279
954	GGTCATCACCA-----	964
279	roPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAla	299
965	-----AGGTGGACAGCGTGTGACCCCGGAGGAGCGCTCAGAGCCAGTCCGCG	1010
299	laProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAla	319
1011	CACGGCATCTCCACGCCCGCAGTGTC-----CAGACGGCTCGCTCGCGGACGA	1061
319	euLeuProLeu-----CysAlaTyrGlyAlaGlyGluProAla	331
1062	CCACGCGCGCGCCCTAAAGGGCTGCCCCGCTTCACGGTGGACACCATCA	1111

```

RESULT 15
US-09-976-594-495
; Sequence 495, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Suchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER C
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 495
; LENGTH: 5080
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 206866.1
; NAME/KEY: unsure
; LOCATION: 2468
; OTHER INFORMATION: a, t, c, g, or other
; US-09-976-594-495

```

Alignment Scores:

Pred. No.: 1,068-08 Length: 5080  
Score: 324.50 Matches: 106  
Percent Similarity: 44.55% Conservative: 29  
Best Local Similarity: 34.98% Mismatches: 88  
Query Match: 15.28% Indels: 81  
DB: 4 Gaps: 11

US-10-087-080-32 (1-402) x US-09-976-594-495 (1-5080)

```
QY 109 ArgSerLysProTyrThrArgArgProLysProProTyrSerTyrIleAlaLeuIleAla 128
DB 23 CQTTCCAGCGCAGCTACCGCAGCCAAAGCGCCCTACTCGTACATCTCGCTCATCACC 82
QY 129 MetAlaIleArgAspSer-AlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLe 148
DB 83 ATGGGCATCCAGCAGCGCCCGCCAGCAAGATGCTCAGCTGAGCGAGATCTACCAGTGGAT 142
QY 148 uMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSerValArgHi 168
DB 143 CATGGACCTCTTCCCTATTACCGGCAGAACCCAGCAGCGCTGGCAGAACTCCATCCGCCA 202
QY 168 sAsnLeuSerLeuAsnAspCysPheValIysValLeuArgAspProSerArgProTyrGl 188
DB 203 CTCGCTGTCTTCAATGACTGCTTCTCAAGTGGCAGCGCTCCCGGACAGCCG---GG 259
QY 188 VlysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPh 208
DB 260 CAAGGCTCTCTACTGACCGCTGCACCGGACTCCGGCAACATGTTCCGAGACGGCTGCTA 319
QY 208 eArgArgArgArgLysArgLysHisArgAlaProValProAlaProGlyLeuArgPr 228
DB 320 CTTGCGCCGC-----CCTCCATCGGGGTGTGCACG 349
QY 228 oGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaAlaProAlaSerPr 248
DB 350 GGAAGACCGCCAGCTAG-AGGCGCGCGCGGCCCGCCCGCG-----CGCCACGCCCC 402
QY 248 oArgMetArgSerProAlaArgGlnGluArgAlaSerProAlaGlyLysPheSerSe 268
DB 403 CAGACTCTGGACCACACAGTGGGGCGGCGGACGAGGGGGCGCT----- 445
QY 268 rSerPheAlaIleAspSerIleLeuArgLysProPheArgSerArgArgLeuArgAspTh 288
DB 446 -----CGAGTTGAGACTCCAGCCTCC 468
QY 288 r-----AlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro---Cy 301
DB 469 TCAACTGGCGCCCGCCATAAGTCCGGGCGCG-----GGCGGCTGGCTCTGTG 516
QY 301 sProProLeu-ProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuP 321
DB 517 CCGCGCTCTCACCCTGC----- 533
QY 321 roLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAla---ArgGluAlaGluV 340
DB 534 -----ACAGGCTTGGCACCCCGCAGTCCCGAGTCCAGCTG 564
QY 340 alProProThrAlaProProLeuLeuAlaPro----- 351
DB 565 CACCTGAAGGGGACCGCCACTACTCTTCAACACCCCGTCTCCATCAACACCTCATG 624
QY 352 --LeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHisL 371
DB 625 TCCTCTCGGAGCAGCAGCAGTACGCTGAGCTTCAAGCATACGACAGGCACTGCAATAC 684
QY 371 euTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyP 391
DB 685 TC-----GCCTTAGGCTTCTACGTTGCCCGCAGCGCTGCTCTAGGCAGCGCTCGGTG 738
QY 391 roHis 392
|||
```

DB 739 ACCAC 743

Search completed: March 31, 2004, 08:40:24  
Job time : 130 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 31, 2004, 08:01:11 ; Search time 396 Seconds  
(without alignments)  
3779.419 Million cell updates/sec

Title: US-10-087-080-32  
Perfect score: 2123  
Sequence: 1 MKLEVFVRAAHGDKQGSDDL.....AALVRRPGPHLSYPVETILA 402

Scoring table:  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2459946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DBV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10087080/runat\_29032004\_114948\_24643/app.query.fasta\_1.583  
-DB=Published Applications\_NA -QWTF=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATH=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10087080 @CIGN 1 1 107 @runat\_29032004\_114948\_24643  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Match Length DB ID Description  
-----  
Pred. No. : 3.79e-142 Length: 1209

1	2123	100.0	1209	12	US-10-229-345-9	Sequence 9, Appli
2	2123	100.0	1209	12	US-10-274-177-9	Sequence 9, Appli
3	2123	100.0	1209	15	US-10-087-080-31	Sequence 31, Appl
4	716	33.7	585	15	US-10-027-632-231353	Sequence 231353
5	488.5	23.0	2487	14	US-10-205-823-135	Sequence 135, App
6	487.5	23.0	2271	14	US-10-101-510-133	Sequence 133, App
7	470	22.1	320	9	US-09-833-381-1266	Sequence 1266, Ap
8	466.5	22.0	2187	10	US-09-960-706-947	Sequence 947, App
9	423	19.9	1662	10	US-09-292-862-1	Sequence 1, Appli
10	419.5	19.8	2753	14	US-10-007-280A-115	Sequence 115, App
11	411	19.4	2593	14	US-10-293-582-24	Sequence 24, Appl
12	411	19.4	3098	14	US-10-198-846-10858	Sequence 10858, A
13	407.5	19.2	1137	9	US-10-029-386-24788	Sequence 24788, A
14	406	19.1	2872	14	US-09-969-708-455	Sequence 455, App
15	406	19.1	2872	14	US-10-177-293-211	Sequence 211, App
16	405	19.1	3289	9	US-09-963-285-8	Sequence 8, Appli
17	405	19.1	6458	9	US-09-963-285-1	Sequence 1, Appli
18	403	19.0	2712	9	US-09-963-285-6	Sequence 6, Appli
19	403	19.0	6021	9	US-09-963-285-5	Sequence 5, Appli
20	391.5	18.4	1506	9	US-09-963-285-9	Sequence 9, Appli
21	391	18.4	4454	9	US-09-764-887-496	Sequence 496, App
22	391	18.4	4454	14	US-10-073-961-496	Sequence 497, App
23	387	18.2	4450	14	US-10-073-961-497	Sequence 497, App
24	387	18.2	4450	14	US-10-029-386-20688	Sequence 20688, A
25	358.5	16.9	7114	14	US-10-388-934-44	Sequence 44, Appl
26	349	16.4	2218	15	US-10-029-386-10008	Sequence 10008, A
27	346	16.3	599	14	US-09-867-701-10873	Sequence 10873, A
28	327	15.4	3119	9	US-09-918-995-4735	Sequence 4735, Ap
29	325.5	15.3	477	10	US-10-027-632-138700	Sequence 138700,
30	314	14.8	425	15	US-10-027-632-138701	Sequence 138701,
31	314	14.8	425	15	US-09-963-285-3	Sequence 3, Appli
32	313.5	14.8	4158	9	US-10-027-632-138703	Sequence 138703,
33	313	14.7	425	15	US-10-027-632-138699	Sequence 138699,
34	310	14.6	1605	13	US-10-044-442-10	Sequence 10, Appl
35	309.5	14.6	1605	13	US-09-910-943-448	Sequence 448, App
36	308	14.5	730	9	US-10-044-442-13	Sequence 13, Appl
37	307	14.5	1634	13	US-10-029-386-23708	Sequence 23708, A
38	297	14.0	1668	13	US-10-044-442-16	Sequence 419, App
39	295.5	13.9	312	14	US-10-101-510-419	Sequence 6254, Ap
40	277	13.0	1056	14	US-09-960-352-6254	Sequence 108, App
41	273	12.9	441	9	US-10-007-926A-408	Sequence 134, App
42	264	12.4	3059	14	US-10-007-926A-134	Sequence 52, Appl
43	263	12.4	5723	14	US-10-341-434-52	Sequence 2145, Ap
44	263	12.4	5723	15	US-10-062-674-2145	
45	260	12.2	3029	15		

ALIGNMENTS

RESULT 1  
US-10-229-345-9  
; Sequence 9, Application US/10229345  
; Publication No. US2004038220A1  
; GENERAL INFORMATION:  
; APPLICANT: MARKOWITZ, Sanford D.  
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS  
; FILE REFERENCE: CWRU-P01-003  
; CURRENT APPLICATION NUMBER: US/10/229,345  
; CURRENT FILING DATE: 2002-08-26  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 9  
; LENGTH: 1209  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1161)..(1161)  
; OTHER INFORMATION: n=a, c, g, or t  
US-10-229-345-9

Alignment Scores: 3.79e-142 Length: 1209  
Pred. No.: 3.79e-142 Length: 1209



QY 101 GlyAlaGlySerGlyGluGlyValAlaArgSerLysProTyrThrArgArgProLysProPro 120  
Db 301 GCGCGGGGAGCGCGAGGGTGCACGACGAAGCATATACGCGCGCGCCCAAGCCCCC 360  
QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140  
Db 361 TACTCGTACATCGCGCTCATCGCATGGCCATCCGCACTCGGGGGCGGCGCTTGACG 420  
QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160  
Db 421 CTGGCGGAGATCAACGAGTACCTCATGGCAAGTTCCCTTTTCCGCGGCGAGCTACAG 480  
QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAspCysPheValIysValLeu 180  
Db 481 GGCTGGCGCAACTCGGTGCGCCACAACTTTCGCTCAACGACTGCTTCTGTCAGGGTGTG 540  
QY 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200  
Db 541 CGCGACCCCTCGCGCCCTGGGCGAAGGACAACTACTGATGCTCAACCCCAACAGCGAG 600  
QY 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysArgLeuSerHisArgAlaPro 220  
Db 601 TACACCTTCGCGAGCGGGTCTTTCGCGCGCGCCGCAAGCGCTCAGCCACCGCGCGCG 660  
QY 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240  
Db 661 GTCCCGCGCCCGGCTGCGGCGCGAGAGGCCCCGCGGCTCCCGCGCGCCCGCGGCC 720  
QY 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAla 260  
Db 721 GCGCGCGCGCGCCCGCTGCGCGCGCTGCGCTGCGCGCGCGCGCGCGCGAGGCGCGCC 780  
QY 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280  
Db 781 AGCCCGCGCGGCAAGTCTCCAGCTCTTCGCACTCGACAGCATCTCGCAAGCCCTTC 840  
QY 281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro 300  
Db 841 CGCAGCGCTGCTCAGGGACACGCGCCCGCGGACGAGCTTCAGTGGGGCGCGCGGCC 900  
QY 301 CysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeu 320  
Db 901 TCGCGCGCGCTGCGCGGCTTCCCGCGCTTCCCGCGGCGCGCGCGCGCGCGCGCTG 960  
QY 321 ProLeuCysAlaTyrGlyAlaGlyGluProAlaAlaArgLeuGlyValAlaArgGluAlaGluVal 340  
Db 961 CGGCTCTGCGGTACGCGCGCGGAGCGCGCGCGCTGGGCGCGCGAGCGCGAGGTG 1020  
QY 341 ProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysPro 360  
Db 1021 CCACCGACCGCGCGCGCTCTGCTTGACCTCTCCCGCGCGCGCGCGCGCGCGCGCG 1080  
QY 361 LeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAla 380  
Db 1081 CTCGAGAGCG 1140  
QY 381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThrLeu 400  
Db 1141 CTGACGCGCGCTTAGTCCGNCCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1200  
QY 401 LeuAla 402  
Db 1201 CTAGCT 1206

## RESULT 3

US-10-087-080-31  
; Sequence 31, Application US/10087080  
; Publication No. US20030235820A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Markowitz, Sanford David  
; APPLICANT: Eos Biotechnology, Inc.  
; APPLICANT: Case Western Reserve University

; TITLE OF INVENTION: No. US20030235820A1 Methods of Diagnosis of Metastatic Colorect.  
; TITLE OF INVENTION: Cancer, Compositions and Methods of Screening for  
; TITLE OF INVENTION: Modulators of Metastatic Colorectal Cancer  
; FILE REFERENCE: 018501-000840US  
; CURRENT APPLICATION NUMBER: US/10/087,080  
; CURRENT FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: US 60/272,206  
; PRIOR FILING DATE: 2001-02-27  
; PRIOR APPLICATION NUMBER: US 60/281,149  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: US 60/284,555  
; PRIOR FILING DATE: 2001-04-17  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 31  
; LENGTH: 1209  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: winged helix/forkhead transcription factor (HPH1)  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1161)  
; OTHER INFORMATION: n = g, a, c or t  
; US-10-087-080-31

## Alignment Scores:

Pred. No.: 3 79e-142 Length: 1209  
Score: 2123.00 Matches: 402  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-087-080-32 (1-402) X US-10-087-080-31 (1-1209)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
Db 1 ATGAAGTTGAGGTGTTCCTCCCTCGCGCGCGCCACCGGGGACAAAGCGGACGTGACCTG 60  
QY 21 GUGUGAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40  
Db 61 GAGGGCGCGGGCGGACGACGCGCTCCCTCCCTGTCGCGCGCGGAGACGACTCCCTG 120  
QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60  
Db 121 GGCTCAGATGGGACTCGCGCGCGCAAGCCCTCCCGCGCGCGCGCGCGCGAGATACGAG 180  
QY 61 GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluAlaIleProAlaAla 80  
Db 181 GCGACCGCGCAACAGAGTGGGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 240  
QY 81 AlaAlaAlaValValAlaGluGlyAlaGluGlyAlaGluAlaGlyProGlyAlaGly 100  
Db 241 GCTCTCAGCGGTGTGTGGCGGAGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 300  
QY 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120  
Db 301 GCGCGGGGAGCGCGAGGTGACCGCGCAAGCCATATACGCGCGCGCGCGCGCGCGCC 360  
QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140  
Db 361 TACTCGTACATCGCGCTCATCGCATGGCCATCGCGACTCGCGCGCGCGCGCGCTTGACG 420  
QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160  
Db 421 CTGGCGGAGATCAACGAGTACCTCATGGCAAGTTCCCTTTTCCGCGGCGAGCTACAG 480  
QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeu 180  
Db 481 GGCTGGCGCAACTCGGTGCGCGCGCAAACTTTCGCTCAACGACTGCTTCTGTCAGGGTGTG 540  
QY 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200

Db 541 CGGACCCCTCGGGCCCTGGGCAAGGCAACTACTGATGCTCAACCCCAACAGCGAG 600  
Qy 201 TTTTThPheAlaAspGlyValPheArgArgArgArgArgArgArgArgArgArgArg 220  
Db 601 TACACCTTCGCGCGGGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
Qy 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaProProPro 240  
Db 661 GTCCCGCGCGCGCGGTGGCGCGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 720  
Qy 241 AlaProAlaProAlaProAlaProArgMetArgSerProAlaArgGlnGluArgAla 260  
Db 721 CG 780  
Qy 261 SerProAlaGlyLysPheSerSerSerPheAlaAlaAspSerIleLeuArgLysProPhe 280  
Db 781 AGCCCGCGCGCGAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTTCGCCAAGCCCTTC 840  
Qy 281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro 300  
Db 841 CGCAGCGCTCGCTCAGGACACGGCCCCCGGACGAGCTTCAGTGGGGCGCGCGCGCG 900  
Qy 301 CysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeu 320  
Db 901 TGCCCGCGCGTCCCGCGCTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCG 960  
Qy 321 ProLeuCysAlaTyGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluVal 340  
Db 961 CCGCTCTGCGCGTACGGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020  
Qy 341 ProProThrAlaProProLeuLeuAlaProLeuProAlaAlaProAlaAlaProAlaLysPro 360  
Db 1021 CCACCGACCG 1080  
Qy 361 LeuArgGlyProAlaAlaGlyAlaAlaHisLeuTyrcysProLeuArgLeuProAlaAla 380  
Db 1081 CTCCGAGCG 1140  
Qy 381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrcysProValGluThrLeu 400  
Db 1141 CTGAGCGCGCGCTAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200  
Qy 401 LeuAla 402  
Db 1201 CTAGCT 1206

RESULT 4  
US-10-027-632-231353/c  
; Sequence 231353, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; POLYMORPHISMS IN THE HUMAN GENOME  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 231353  
; LENGTH: 585  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-231353  
Alignment Scores:  
Pred. No.: 1,2e-42 Length: 585  
Matches: 136  
Conservative: 0  
Percent Similarity: 97.84%  
Best Local Similarity: 97.84%  
Mismatch: 3  
Query Match: 33.73%  
Indels: 0  
Gaps: 15  
DB: 0  
US-10-087-080-32 (1-402) x US-10-027-632-231353 (1-585)  
Qy 264 GlyLysPheSerSerSerPheAlaAlaAspSerIleLeuArgLysProPheArgSerArg 283  
Db 583 GGCAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTTCGCCAAGCCCTTCGCCAGCGCG 524  
Qy 284 ArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaProCysProPro 303  
Db 523 CGCTCAGGACACGGCCCCCGGACGAGCTTCAGTGGGGCGCGCGCGCGCGCGCGCG 464  
Qy 304 LeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuProLeuCys 323  
Db 463 CTGCGCGCGTTCGCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 404  
Qy 324 AlaTyrcysAlaTyGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValPro 343  
Db 403 GCGTACG 344  
Qy 344 AlaProProLeuLeuAlaProLeuProAlaAlaProAlaAlaProAlaLysProLeuArgGly 363  
Db 343 GCGCGCGCGCTCTGCTTCACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 284  
Qy 364 ProAlaAlaGlyGlyAlaHisLeuTyrcysProLeuArgLeuProAlaAlaLeuGlnAla 383  
Db 283 CCG 224  
Qy 384 AlaLeuValArgArgProGlyProHisLeuSerTyrcysProValGluThrLeuLeuAla 402  
Db 223 GCCTCAGTCCGCGCGCGCTTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 167  
RESULT 5  
US-10-205-823-135  
; Sequence 135, Application US/10205823  
; Publication No. US20030108963A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Gorbacheva, Bella  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Womsey, Angela M.  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Anderson, Dustin  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS AND  
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; THERAPY OF PROSTATE CANCER  
; FILE REFERENCE: MRI-044  
; CURRENT APPLICATION NUMBER: US/10/205,823  
; CURRENT FILING DATE: 2002-07-25  
; PRIOR APPLICATION NUMBER: 60/307,982  
; PRIOR FILING DATE: 2001-07-25  
; PRIOR APPLICATION NUMBER: 60/314,356  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/325,020  
; PRIOR FILING DATE: 2001-09-25

256 nGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLe 276  
1239 GCGGCTACGCGCTGCGAGCTGCC...GCTTACGCGCGCGCTCGGCTCTTCG 1289  
276 uArgLysProPheArgSerArgArg...LeuArgAspThrAlaPr 290  
1290 CCGCGCAGCGCGCGCGCGCGCGCTTCCACCGCAGCTCGCGCGCGCGCG 1349  
290 oGlyThrLeuGlnTrpGlyAlaAlaProCysProPro...LeuProAlaPh 307  
1350 CGCACCGCAGCGCG...GCGCGCGAGCTGCGCGGAGCGCTTCGGGTACGCGCG 1406  
307 eProAlaLeuLeuPro...AlaAlaProCysArgAlaLeuLeuProLeuCysAlaTy 325  
1407 ACCGCTCG-GCGCGCGCTTACCGCGCGCTTCCGCGCTTCCGCGCGCGCG 1463  
325 rGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAla... 338  
1464 GCGCGCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCG 1519  
339 GCGCGCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCG 1579  
1520 GCGAGCTTGGCGCGCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCG 1579  
350 aProLeuProAlaAlaAlaPro...AlaLysProLeuArgGlyProAlaAlaGly 368  
1580 GCGTTCG-CCTCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCG 1638  
368 yAlaHisLeuTyrcysProLeuArgLeuProAlaAlaLeuGlnAlaLeu-ValArgA 388  
1639 C...TGGCGCGCGCGCGCGCTTCCGCGCTTCCGCGCTTCCGCGCGCG 1668  
388 rgPro...GlyProHisLeuSerTyrcysPro 396  
1669 GCGCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCGCTTCCGCG 1697

RESULT 6  
US-10-101-510-133  
; Sequence 133, Application US/10101510  
; Publication No. US20030148295A1  
; GENERAL INFORMATION:  
; APPLICANT: WAN, JACKSON  
; APPLICANT: WANG, YIXIN  
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE  
; FILE REFERENCE: 15117.0012  
; CURRENT APPLICATION NUMBER: US/10101510  
; CURRENT FILING DATE: 2002-03-20  
; PRIOR APPLICATION NUMBER: 60/276,947  
; PRIOR FILING DATE: 2001-03-20  
; NUMBER OF SEQ ID NOS: 805  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 133  
; LENGTH: 2271  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-101-510-133

Alignment Scores:  
Pred. No.: 6,45e-26 Length: 2271  
Score: 487.50 Matches: 177  
Percent Similarity: 44.01% Conservative: 36  
Best Local Similarity: 36.57% Mismatches: 137  
Query Match: 22.96% Indels: 139  
DB: 14 Gaps: 21

US-10-087-080-32 (1-402) x US-10-101-510-133 (1-2271)

QY 28 AlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAspGly----- 44  
Db 64 GCGCGCGCGCGCG-CCTCGCGCGCGCGAGACCGCGCTTGGCGCGCGAGGAGCGCGGAGC 122  
QY 45 -----AspCysAlaAlaLysPro----- 50

PRIOR APPLICATION NUMBER: 60/341,746  
PRIOR FILING DATE: 2001-12-12  
PRIOR APPLICATION NUMBER: 60/362,158  
PRIOR FILING DATE: 2002-03-05  
NUMBER OF SEQ ID NOS: 455  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 135  
LENGTH: 2487  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-205-823-135

Alignment Scores:  
Pred. No.: 5.94e-26 Length: 2487  
Score: 488.50 Matches: 170  
Percent Similarity: 47.56% Conservative: 35  
Best Local Similarity: 39.44% Mismatches: 147  
Query Match: 23.01% Indels: 83  
DB: 14 Gaps: 20

US-10-087-080-32 (1-402) x US-10-205-823-135 (1-2487)

QY 14 AspLysGlnLysSerAspLeuGluGlyAlaGlyGlySerAspAlaProSerProLeuSer 33  
Db 506 GAAGAGGAGGACGACGAGGCGCGCTGCGCGCGC-CGCGCTGGCTTCCCGCGCA 564  
QY 34 AlaAlaGlyAspSerLeuGlySerAspGlyAspCysAlaAlaLysProSerAlaGly 53  
Db 565 GCGCGCGCGCGCGCGCTGCTGACCGCGGAGAGCGAGCTGGAGGATCTGGAGGAGA 624  
QY 54 GlyGlyAlaArg-AspThrGlnGlyAspGlyGlnSerAlaGlyGlyGlyProGlyAl 73  
Db 625 GGAGGACGAGGATGACATCTG-----CTGGCGCGCGCTTCCGCGCGCG 678  
QY 73 aGluGluAlaIleProAlaAlaAlaAlaAlaValAlaAlaGluGlyAlaGluAlaG 93  
Db 679 CCGCGCGCGCGCGCGCGCGCGCGCA-----GGAGCGCGGTGGGG 720  
QY 93 yAlaAlaGlyProGlyAlaGlyAlaGlySer---GlyGluGlyAlaArgSerTyrcys 112  
Db 721 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGCGCGGTAGCGCGCGCAAGAAC--CC 777  
QY 112 oTyrcysArgProGlyProProTyrcysTyrcysTyrcysTyrcysTyrcysTyrcys 132  
Db 778 GCTGGGT-----AAGCGCGCTTCTGATATCGCGCTTCTGCGCTTCTGCGCTTCT 828  
QY 132 gAspSerAlaGlyGlyArgGluThrLeuAlaGluIleAsnGluTyrcysLeuMetGlyLys 152  
Db 829 GCAGAGCCCAAGAGCGCGCTGAGCTGAGCGAGATCTGTGATTCATCAGCGCGCTT 888  
QY 152 eProPhePheArgGlySerTyrcysTyrcysTyrcysTyrcysTyrcysTyrcys 172  
Db 889 CCCTACTACCGGAGAGTTCCTCGCTGCGAGACAGCATCCGCGCAAGCTCTCGT 948  
QY 172 uAsnAspCysPheValLysValLeuArgAspProSerArgProTyrcysTyrcysTyrcys 192  
Db 949 CAACGAGTCTTCTGTAAGATCCCGCGAGCGCGCGCGCAACCCCGCGCAAGCTCTCT 1005  
QY 192 rTrpMetLeuAsnProAsnSerGluTyrcysTyrcysTyrcysTyrcysTyrcysTyrcys 212  
Db 1006 CTGGACCTGACCGCGAGTCCCGCAGCATGTTTCGACACCGCGAGCTTCTGCGCGGAG 1065  
QY 212 gLysArgLeuSerHisArgAlaProValProAlaProGlyLeuArgProGluGluAla-- 231  
Db 1066 GAAGCGCTTCAAG---CGGAGCGCGTGTCTCCACCGCGCGCGCGCGCGCGGTCTCT 1122  
QY 232 -----ProGlyLeuProAlaAla----- 237  
Db 1123 GCTCTCGCT 1182  
QY 238 -----ProProProAlaProAlaProAlaProAlaProAlaProAlaProAlaPro 256  
Db 1183 CTTCCGCGCGCGC-GCT 1238

```

Db 123 CCGGCGCCCATGCGCGCCCGCCCGCCCGCCAGCTATGACCCCTGAGCACTGA 182
Qy 51 -----SerAlaaglyGlyGlyValaArg 57
Db 183 GATGTCGATGCTCTGCGCTCGCGAGGAACAGACATCGACGTGTGGGGAGGCGA 242
Qy 58 -AspThrIngly-----AspGlyGluInSerAlaGlyGlyProGlyAl 73
Db 243 GACGAGAGAGACGAGAGAGAGAGACACAGCAGGCGCGGTGGCGGCCCG-GC 301
Qy 73 aGluGluAlaIleProAlaAlaAlaAlaValValaGluGluAlaGluAlaGl 93
Db 302 TGGCTGTCCCGCGCAGCGCGCGCGCGCGCGCTCGTACCGCGGGAGACAGCTGG 361
Qy 93 Y----- 93
Db 362 AGGATCTGAGGAGGAGGAGGAGGAGCATGACATCTGCTGCGCCCGCTCTGGGGCTC 421
Qy 94 -----AlaAlaGlyProGlyAlaGlyGlyAlaGlySerGl 105
Db 422 CCGGCGCGCCCGCGCGCCCGCGCGCGCGCGCGAGCGCGTGGGGCGCGCGCG 481
Qy 105 yGluGlyAla-----ArgSerLysProTyrThrArgArgPro-----LysPr 119
Db 482 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 541
Qy 119 oProTyrSerTyrIleAlaAlaIleAlaMetAlaIleArgAaspSerAlaGlyArgLe 139
Db 542 GCCTACTCTGTATATCGCTCTATCATCTATGCTCTCTGAGAGAGAGAGAGCGCT 601
Qy 139 uThrLeuAlaGluIleAasnGluTyrLeuMetGlyLysPheProPheAsgGlySerTy 159
Db 602 GACGCTGAGCGAGATCTGTGAGTTCATCAGCGCGCGCTTCCCTACTACCGGAGAGTT 661
Qy 159 rThGlyTyrPheAaspSerValArgHisasnLeuSerLeuAasnAspCysPheValLysVa 179
Db 662 CCGCGCTGTCGACAAACAGATCGCCACCACTCTCGCTCAACGACTGCTTCGCAAGAT 721
Qy 179 lLeuArgAaspProSerArgProTyrGlyLysAasnTyrTrpMetLeuAasnProAasnSe 199
Db 722 CCGCGCGAGCGCGCAACCG-GGCAGGGGCACTACTGGAGCTGGACCGGAGTC 778
Qy 199 rGluTyrThrPheAlaAaspGlyValPheArgArgArgArgLysArgLeuSerHisArgAl 219
Db 779 CGCGCATGTTTCGACAAACGCGAGCTTCTGCGCGCGAGGAGCGCTTCAAG-CCGCA 835
Qy 219 aProValProAlaProGlyLeuArgProGluGluAla----- 231
Db 836 GCGGCTGCTCCACCAACCGCGCGCGCGCGCGAGTCTGTGCTGCGCGCGCGGAGC 895
Qy 232 -----ProGlyLeuProAlaAla-----ProProProAlaProAl 243
Db 896 CGACGGGGCGCGGGGAGCGCGGACCGCGCGCGCTCTTCGCGCGCGC-GCCCGCGC 954
Qy 243 aAlaProAlaSerProArgMetArgSerProAlaArgGluGluArgAlaSerProAl 263
Db 955 CGCCCCCGCATGCTACGCTACGG-CCCTACGGCTGGCGGTACGCGCTGCAGCTGC 1011
Qy 263 aGlyLysPheSerSerPheAlaIleAaspSerIleLeuArgLysProPheArgSerAr 283
Db 1012 C-----GCCTTACGCGCGCGCTCGCGCTCTTCGCGCGGACGCGCGCGCGCG 1062
Qy 283 gArg-----LeuArgAaspThrAlaProGlyThrThrLeuGlnTyrGl 297
Db 1063 CGCGCGCGGCTTCCACCGCACTCGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
Qy 297 yAlaAlaProCysProPro-----LeuProAlaPheProAlaLeuLeuPro----- 312
Db 1120 CCGCGAGCTGGCGCGGACCGGCTTGGCTACCGCGCGCGCGCGCGCGCGCGCGCTA 1178
Qy 313 -AlaAlaProCysArgAlaAlaLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaAr 332

```

```

Db 1179 CCGGCGCCCTGCTCGCGCTCGCGCGCCNAGCGGGGG-CCCGCGGGGCGCTCAGCG 1232
Qy 332 gLeuGlyAlaArgGluAla----- 338
Db 1233 CTGGCGCGCTCGCCCTTCTCCATCGAGAGCATCATCGGGGCGAGCTTGGGCGCGCGCC 1292
Qy 339 ----GluValProProThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaPr 357
Db 1293 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAGCGCTCG-CCCTCGCGCTCGCC 1351
Qy 357 o-----AlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLe 375
Db 1352 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAGCGAGGCG- 1391
Qy 375 uArgLeuProAlaAlaLeuGlnAlaAlaLeu-ValArgArgPro---GlyProHisLeuS 394
Db 1392 -----TCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTCACCAGATCCCT 1441
Qy 394 eTyrPro 396
Db 1442 CGTGGCG 1449

```

RESULT 7  
 US-09-833-381-1266  
 ; Sequence 1266 Application US/09833381  
 ; Patent No. US20020132090A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Robison, Keith E.  
 ; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs  
 ; FILE REFERENCE: 5800-119  
 ; CURRENT APPLICATION NUMBER: US/09/833,381  
 ; PRIORITY FILING DATE: 2001-04-11  
 ; PRIOR APPLICATION NUMBER: 09/516,448  
 ; PRIORITY FILING DATE: 2000-02-29  
 ; NUMBER OF SEQ ID NOS: 2050  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 1266  
 ; LENGTH: 320  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(320)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-833-381-1266

Alignment Scores:  
 Pred. No.: 1,96e-25 Length: 320  
 Score: 470.00 Matches: 94  
 Percent Similarity: 89.62% Conservative: 1  
 Best Local Similarity: 88.88% Mismatches: 2  
 Query Match: 22.14% Indels: 9  
 Gaps: 0

US-10-087-080-32 (1-402) x US-09-833-381-1266 (1-320)  
 Qy 190 AspAsnTyrTrpMetLeuAasnProAasnSerGluTyrThrPheAlaAaspGlyValPheArg 209  
 Db 2 GACAACTACTGATGCTCAACCCCAACAGCGAGTACACCTTCGCGCGAGCGGGTCTCCCG 61  
 Qy 210 ArgArgArgLysArgLeuSerHisArgAlaProValProAlaProGlyLeuArgProGlu 229  
 Db 62 CGCGCGCGCAAGCGCTCAACCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCG 121  
 Qy 230 GluAlaProGlyLeuProAlaAlaProProProAlaProAlaProAlaSerProAr 249  
 Db 122 GAGGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181  
 Qy 249 gMetArgSerProAlaArgGluGluArgAlaSerProAlaGlyLysPheSerSerse 269  
 Db 182 CATGCGCTCG 241  
 Qy 269 rPheAla-IleAaspSerIleLeuArgLysProPheArgSerArgLeuArgAspThrA 289

Db 242 CTTTCATATCTACAGCATCTGCNCAATCCCTTCGCGATCCGCGCGCTCANGGACACGG 301  
Qy 289 laProGlyThrThr 293  
Db 302 CCCCCGGGAGACC 315

## RESULT 8

US-09-960-706-947  
; Sequence 947, Application US/09960706  
; Publication No. US20030134280A1  
; GENERAL INFORMATION:  
; APPLICANT: Munger, William E.  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia  
; FILE OF INVENTION: Gene Expression Profiles  
; FILE REFERENCE: 44921-5029-0105  
; CURRENT APPLICATION NUMBER: US/09/960,706  
; CURRENT FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: 60/223,323  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: 09/873,319  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 947  
; LENGTH: 2187  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 U13220  
US-09-960-706-947

Alignment Scores:  
Pred. No.: 1,936-24 Length: 2187  
Score: 456.50 Matches: 157  
Percent Similarity: 45.35% Conservative: 48  
Best Local Similarity: 34.73% Mismatches: 159  
Query Match: 21.97% Indels: 90  
DB: 10 Gaps: 18

US-10-087-080-32 (1-402) x US-09-960-706-947 (1-2187)

Qy 8 ProArgAlaAlaHisGlyAspGlyGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27  
Db 79 CGCCCTCCGCGCCGCGGCTCCTCCGCGTCCAGATGACACCGAGCGGGG---CCG 135  
Qy 28 AlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySer---AspGlyAspCys 46  
Db 136 CGCGCGCGCGCGTCCGCGCGCGTGCAGCCCGTCCCGCGCGCTCCAGCGCCCTCG 195  
Qy 47 AlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAspGlyGlnSer 66  
Db 196 ATGAGCCCG 255  
Qy 67 AlaGlyGlyGlyProGlyAlaGluGluAlaProAlaAlaAlaAlaValVal 86  
Db 256 TCTCTGCTGCT 315  
Qy 87 AlaGluGlyAlaGluAlaGlyAlaGlyProGlyAlaGlyAlaGlyAlaGlyGly 106  
Db 316 GCGCCCTCGCT 375  
Qy 107 GlyAlaArgSerLysProThrArgArgPro---LysProProThrSerTyrIleAla 125  
Db 376 GCCAAGAGCGGAGCTCGGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 435  
Qy 126 LeuIleAlaMetAlaIleArgAspSerAlaGlyArgLeuThrLeuAlaGluLeuAla 145  
Db 436 CTCATGTCATGGCCATCCAGAGCTCGCCCAAGCGGCTCAGCTCAGCGAGATCTAC 495  
Qy 146 GluTyrLeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSer 165  
Db 496 CAGTCTCTGAGCGCGCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 555

Qy 166 ValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeuArgAspProSerArg 185  
Db 556 GTGCGCCACAATCTCTCGCTCAACGAGTGTCTTCATCAAGCTGCTTAAGGCGCTCGGGCGG 615  
Qy 186 ProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAsp 205  
Db 616 CCC---GGCAGGGCCACTACTGACCATCGACCGCGCGCGCGCGCTTCATGTTGCGAGGAG 672  
Qy 206 GlyValPheArgArgArgLysArgLeuSerHisArgAlaPro----- 220  
Db 673 GGCTCGTTCGCGCGCGGCC-GCGCGGCTTCAGCGGGAAGTGCAGCGCGCTCAAGCCCAT 731  
Qy 221 ValPro----- 227  
Db 732 GTACCAACCGCGTGTGAGCGGCTTGGCTTCGGGGGCTCGCTGCTCCCGAGGGCTTCGA 791  
Qy 228 ProGluGluAlaProGlyLeuProAlaAlaProProProAlaProAlaProAlaSer 247  
Db 792 -----CTTCAGGCGCGCGCGCTCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCG 833  
Qy 248 ProArgMetArgSerPro-----AlaArgGln 256  
Db 834 GGGCGGCTACGCGCGCTTCGACATGATGCCCGGGGTAGCAGCGCGCGCGCGCGCGCGCG 893  
Qy 257 GluGluArgAla-----SerProAlaGlyLysPheSerSer 268  
Db 894 CAGCCAGCGCACCTCACCACCAACACCAACACCAACACCAACACCAACACCAACACCC 953  
Qy 269 SerPheAlaIleAspSerIleLeu-----ArgLysProPhe 280  
Db 954 GGGTTCACCTACATGCGCGAGCTGCCGGTGCCTGGCGGGGCGCGCGCGCGCGCGCGCG 1013  
Qy 281 ArgSerArgArgLeuArg-AspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPr 300  
Db 1014 CGGGGGCG 1061  
Qy 300 cYsProProLeuProAlaPhe-----ProAlaLeuLeuPr 312  
Db 1062 GTATACCTCGTCCCGCGCATGCGGAGCGCATCGAATGCCACTCGCGCTTACAGAGCCC 1121  
Qy 312 oAlaAla-----ProCysArgAlaLeuLeuPro-----LeuCysAlaTy 325  
Db 1122 TGGGCGCATGAGAGCTGCGCTGG-CGCTGCGCTTACCTCAAGCAGCGCGCTGCGCTGA 1180  
Qy 325 rGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProThrAlaPr 345  
Db 1181 CGCCAGCAGCAACCGCGCGCTCGCGAGG-----CCTGCACTCCAGCA 1225  
Qy 345 oProLeuLeuAlaProLeuProAlaAlaAlaProAlaAlaLysProLeuArgGlyProAl 365  
Db 1226 TGCTCTCTACTCGTGGAGCAGAGCTTGTGACCAAGACGCTCG---CGAGGACCTCT 1282  
Qy 365 aAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaLe 385  
Db 1283 CAGTGAGCATGCC-----CGTTACCAAGCATCACTCTACTCTCAATG 1324  
Qy 385 u---ValArgArgProGlyProHisLeuSerTyr 395  
Db 1325 GTGACAGAAAAGATTGCTCTCACTCAATG 1358

## RESULT 9

US-09-232-862-1  
; Sequence 1, Application US/09292862  
; Publication No. US20030013087A1  
; GENERAL INFORMATION:  
; APPLICANT: Walter, Michael A.  
; APPLICANT: Jordan, Tim  
; APPLICANT: Raymond, Vincent  
; TITLE OF INVENTION: NOVEL MUTATIONS IN THE FREAC3 GENE FOR  
; TITLE OF INVENTION: DIAGNOSIS AND PROGNOSIS OF GLAUCOMA AND ANTERIOR SEGMENT  
; TITLE OF INVENTION: DYSENESIS  
; FILE REFERENCE: 07540/020003

```

/ CURRENT APPLICATION NUMBER: US/09/292,862
/ CURRENT FILING DATE: 1999-04-16
/ EARLIER APPLICATION NUMBER: 60/084,784
/ EARLIER FILING DATE: 1998-05-08
/ EARLIER APPLICATION NUMBER: 60/082,206
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: FASTSEQ for Windows Version 3.0
/ SEQ ID NO 1
/ LENGTH: 1662
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-292-862-1

```

```
US-10-087-080-32 (1-402) x US-10-007-280A-115 (1-2753)
QY 15 LysGlnGlySerAspLeuGluGlyAlaGlyGly-----25
Db 84 CGACAGGCTCCAGCAAGAACGCGCGGCATCCCGAGTCTCCAGAAATTTGAGACT 143
QY 26 -----SerAspAlaProSerPro-LeuSerAlaAlaGlyAspAspSerLeuGlyse 42
Db 144 TGGCCGTAAGCGAGTCTGTCGCGCCCAACTCTTTCCGCGCGCCAGCGCTTGAGGGAGAG 203
QY 42 rAspGlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAs 62
Db 204 CAGAGCGCGCGCGCGCG-----GCGCGCGCGCTTTGTCATGA 242
QY 62 pGlyGlu-----GlnSerAlaGlyGlyProGlyAlaGluGluAlaIlePr 78
Db 243 TGGCCAGTACCCGAGCGCGAGCGCGCGCGCG-----CC 281
QY 78 oAlaAlaAlaAlaAlaValVaAlaGluGlyAlaGluAlaGlyAlaAla-GlyProG 98
Db 282 TGTGCGCCCGAGACCGCTGCACAGTCAAGAGGCCAGAGGCGCGCGCGCGCGAG 341
QY 98 lYAlaGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProL 118
Db 342 GCAAGCGCGTGGGGTGGCGCGGAGCAGCCCGGAGAGCGCGCGCGCGAG---A 398
QY 118 ySProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyAla 138
Db 399 AGCCCGCTACTGTAGTGGCGCTCATCGCCATGCGATCGCGAGAGCGCGGAGAAGA 458
QY 138 rGLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlyLys 158
Db 459 GGCTCAGCTGTCCGCGATCTACAGTACATCCGGAAGTTCCGTTCTACGAGAAGA 518
QY 158 exTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValL 178
Db 519 ATAAGAAGGCTGCACAAATAGCATCCGCCACCACTCAGCTCAACAGAGTCTCATCA 578
QY 178 ySValLeuArgAspProSerArgProTyrGlyLysAspSerTyrMetLeuAsnProA 198
Db 579 AGGTGCGCGCGAGCGCGCGCGAG---CGAAGGGCACTACTGAGCGCTGGACCGCG 635
QY 198 snSerGluTyrThrPheAlaAspGlyValPheArgArgArg-----LysArg-Leu 215
Db 636 CTGCGAAGACATGTCGAGAGGGCACTACCGCGCGCGCGCGCGCATGAAGAGCGCT 695
QY 216 SerHisArgAlaPro-----ValProAlaProGlyLeuArgProGluGluAlaPro 232
Db 696 TCGCGCGCGCGCGCGCGCACTTCCAGCGCGCAAGGGGCTCTTCGGGCGCGAGCGCG 755
QY 233 GlyLeuProAlaAla-----237
Db 756 CAGCGGCGTGGCGTGGCGGCGCGCGCGCGCGCGCGCTACGCTACGCTGCGCGCGCG 815
QY 238 -----ProProPro 240
Db 816 AGTACTGTCAGTCTGGCTTCTCAACAACCTCGTGGCGCTTACCGCAGCTCCCTCACCA 875
QY 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArg---GlnGluGluArg 259
Db 876 TGGCCTATGCTCTGCTGC-----AGATGGCGGAGCGCGCGCGCGGTGAGCAGTGGCG 929
QY 260 AlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysPro 279
Db 930 CTGCAGCGCGCGCGCGCGGTAGCC-----CTGGCGCGCGCG 965
QY 280 PheArgSerArgArgLeuArgAspThrAlaProGlyThr-----292
Db 966 CTGTGTCAGAGGCTCGCGCGCGCGCGCGCGCTGTCAGCGCGCGCGTACACAGCGGTGAGA 1025
QY 293 -----Thr-LeuGlnTyrGlyAlaAlaProCysProPr 303
```

```
Db 1026 GCATGGCGTGGCGCGCGCGGTAGTAGTAAGTCTGACATGGCTGGGAGCGCGCGCG 1085
QY 303 oLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuProLeuCy 323
Db 1086 CACCCCGCGCTCGCGCGCACCCCGCATCC-----1119
QY 323 sAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAla-GluValProPro- 342
Db 1120 -----GCAGCGACCATCTGCACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1169
QY 343 -----ThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaProA 358
Db 1170 ACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1229
QY 358 lAlysProLeuArgGlyProAlaAlaGlyAlaHisLeuTyrCysProLeuArgLeuP 378
Db 1230 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1258
QY 378 roAlaAlaLeuGlnAlaAlaLeuValArgArgProGly-----ProHisLeuSerT 395
Db 1259 --CGGCGCTGCGAGTTGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1316
QY 395 Yr 395
Db 1317 AC 1318

RESULT 11
US-10-293-582-24
; Sequence 24, Application US/10293582
; Publication No. US20030175253A1
; GENERAL INFORMATION:
; APPLICANT: Akil, Huda
; APPLICANT: Bunney, William E.
; APPLICANT: Burke, Sharon
; APPLICANT: Choudary, Prabhakara V.
; APPLICANT: Cox, David R.
; APPLICANT: Evans, Simon
; APPLICANT: Jones, Edward G.
; APPLICANT: Li, Jun
; APPLICANT: Lopez, Juan F.
; APPLICANT: The Trustees of The Leland Stanford Junior University
; TITLE OF INVENTION: Compositions and Methods for Diagnosing and Treating
; TITLE OF INVENTION: Mental Disorders
; FILE REFERENCE: 020885-000210US
; CURRENT APPLICATION NUMBER: US/10/293,582
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/339,252
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
; LENGTH: 2559
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human brain factor 1, HBF-1 transcription factor
US-10-293-582-24

Alignment Scores:
Pred. No.: 1,91e-20 Length: 2559
Score: 411.00 Matches: 131
Percent Similarity: 40.46% Conservative: 28
Best Local Similarity: 33.33% Mismatches: 145
Query Match: 19.36% Indels: 89
DB: 14 Gaps: 12

US-10-087-080-32 (1-402) x US-10-293-582-24 (1-2559)
QY 28 AlaProSerProLeuSerAlaAlaGlyAspAspSerLeuGlySerAspGlyAspCysAla 47
Db 419 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 457
QY 48 AlaLysProSer-AlaGlyGlyAlaArgAspThrGlnGlyGluGlnSerAl 67
```





Db 335 GACCACAGTGGGCGGACGCGGACGAGGGGGCGCTCGAGTTCAAGACTCCAGCTCCTCA 276  
 Qy 312 roAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyGlyAlaGlyGluProAlaA 332  
 Db 275 ACTGCCCGCCCCATAAG-----CTCGGCGCGGGCGCTGGCTCTGTGGCC 229  
 Qy 332 rgleuGly-----AlaArgGluAlaGluValProProThralap 345  
 Db 227 GCCTCTCACCGGCGACACGCGCTTGGCACCCACGAGTCCAGCTGCACCTCGAAGGGGAC 168  
 Qy 345 roProLeuLeuAlaPro-----LeuProAlaAlaA 356  
 Db 167 CCCACTACTCTCTCAACACCGCTTCTCCATCAACACCTCATGCTCTCTCGAGGAG 108  
 Qy 356 laProAlaLysProLeuArgGlyProAlaAlaAlaGlyGlyAlaHisLeuTyCysProLeuA 376  
 Db 107 CAGCATAAGCTGACCTTCAAGGCATACGAACAGGCACTGCAATATC-----GCCATTAC 54  
 Qy 376 rgleuProAlaAlaLeuGlnAlaAlaLeuValArgProGlyProHis 392  
 Db 53 GGCTCTACGTTGCCCGCCAGCTGCTCTAGCAGCGGCTCGGTGACCAC 4  
 RESULT 14  
 US-09-969-708-455  
 ; Sequence 455, Application US/09969708  
 ; Patent No. US20020102532A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Augustus, Meena  
 ; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu  
 ; FILE REFERENCE: 689290-70  
 ; CURRENT APPLICATION NUMBER: US/60/237,606  
 ; PRIOR FILING DATE: 2001-10-03  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR FILING DATE: 2000-10-03  
 ; NUMBER OF SEQ ID NOS: 658  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 455  
 ; LENGTH: 2872  
 ; TYPE: DNA  
 ; ORGANISM: Homosapiens  
 US-09-969-708-455  
 Alignment Scores:  
 Pred. No.: 4,79e-20 Length: 2872  
 Score: 406.00 Matches: 142  
 Percent Similarity: 45.64% Conservative: 41  
 Best Local Similarity: 35.41% Mismatches: 146  
 Query Match: 19.12% Indels: 75  
 DB: 9 Gaps: 13  
 US-10-087-080-32 (1-402) x US-09-969-708-455 (1-2872)  
 Qy 24 GlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43  
 Db 349 GGCTCGCGGGCGCCATGACAGCATGACTGGCGCGCGCTGACGGCCATGGTACG--- 405  
 Qy 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGlnGlyAspGly 63  
 Db 406 -----GGCTGAGCCCGAGC---GGCATGGCGCCATGGGTGGCGCAGCGCGGCC 453  
 Qy 64 GluGlnSerAlaGlyGlyProGlyAlaGluGluAlaAlaLeuValArgProGlyProHis 83  
 Db 454 TCCATGATGATGGCTTGGCCCTTACGGCGCGCCATGAACCCGTGTCATGAGCCCCATG 513  
 Qy 84 AlaValAlaGluGluAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyGlyAlaGly 103  
 Db 514 GGTACGCGCGCTTCAACCTGGCGCGCGCAGCGCGCGGCGC-----GGCGGCGCGGAC 564

Qy 104 SerGlyGluGlyAlaArgSerLysProTyThrArgArgProLysProProTySerTy 123  
 Db 565 GCCAAGAGCTTCAAGCGCGATTACCCGACGCC-----AAGCCGCCCTACTCTGTAC 615  
 Qy 124 lleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGlu 143  
 Db 616 ATCTCGCTCATCACCATGGCCATCGAGCGGCGGCCAGCAAGATGCTCAAGCTGAGCGAG 675  
 Qy 144 lleAsnGluTyLeuMetGlyLysPheProPheArgGlySerTyThrGlyTTPArg 163  
 Db 676 ATCTACAGTGGATCATGGACCTTCTCCCTATTATTCGGCAGACACAGCAGCGCTGGCAG 735  
 Qy 164 AsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspPro 183  
 Db 736 AACTCCATCCGCCCATCTCGCTTCTCAATGACTGCTTCTCAAGGTGGCAGCGCTCCCG 795  
 Qy 184 SerArgProTTPGlyLysAspAsnTyTTPMetLeuAsnProAsnSerGluTyThrPhe 203  
 Db 796 GNCAGCGG---GGCAGGGGCTCTACTGGACGCTGCACCGGACTCCGCCAACATGTTT 852  
 Qy 204 AlaAspGlyValPheArgArgArgLysArgLysSerHisArgAlaProValProAla 223  
 Db 853 GAGAACGGCTGCTATTGGCGCGCCAGAGCGC----- 885  
 Qy 224 ProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAla 243  
 Db 886 -----TTCAAGTGGAGAACGACGCGCGGGG---CCGCGCGGGGGCGGAGCGGAGCG 937  
 Qy 244 AlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAlaSerProAla 263  
 Db 938 GGGGCGAGCGGCGCCAAAG---GCGGC-CCTGAGAGCGCGCAAGGACCCCTCT 984  
 Qy 264 GlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPheArgSerArg 283  
 Db 985 GGC-----GCCTCTAACCAGCGCGGACTCGCCCTCCATCGGGGTGTGCAC----- 1032  
 Qy 284 ArgLeuArgAspThrAlaProGlyThrLeuGlnTrpGlyAlaAlaProCysProPro 303  
 Db 1033 -----GGGAAGACCGCGCAGCTAGAGGGCGCGCGCGCGCGCGCGCGC 1071  
 Qy 304 LeuProAlaPheProAlaLeuLeu----- 311  
 Db 1072 CGGCGCGCGACCCCGACACTCTCGACACAGTGGGCGGACGCGACAGGGGCGCGCTCG 1131  
 Qy 312 -----ProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyGlyAla 327  
 Db 1132 GAGTTGAAGACTCCAGCTCTCTCAACTGC-GCCCCCATAAGCTCGGCGCGCGCGCT 1190  
 Qy 328 GlyGluProAlaArgLeuGly-----AlaArgGluAlaGluVal 340  
 Db 1191 GGCCTCTGTGCGCGCTCTACCGCGGACACGCGTGGCACCACCGAGTCCCGAGTGCA 1250  
 Qy 341 ProProThrAlaProProLeuLeuAlaPro----- 351  
 Db 1251 CCTGAAGGGGACCCCGCCTACTCTTCAACACCGCGCTTCCATCAACCTCATGCTGTC 1310  
 Qy 352 LeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHisLeu 371  
 Db 1311 CTCCTCGGAGCAGCAGCAGCATAAGCTGGACTTCAAGGCATACGAACAGGCACTGCAATCTC 1370  
 Qy 372 TyrCysProLeuArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgProGlyPro 391  
 Db 1371 -----GCCTTACGGCTTACGTTGGCGCGCGCGCTTCTAGGCGAGCGCGCTGTC 1424  
 Qy 392 His 392  
 Db 1425 CAC 1427

RESULT 15  
 US-10-177-293-211  
 ; Sequence 211, Application US/10177293  
 ; Publication No. US2003012128A1  
 ; GENERAL INFORMATION:

```

/ APPLICANT: Lillie, James
/ APPLICANT: Glatt, Karen
/ APPLICANT: Zhao, Xumei
/ APPLICANT: Gannavarpu, Manjula
/ APPLICANT: Kamatkar, Shubhangi
/ APPLICANT: Mertens, Maureen
/ APPLICANT: Myer, Vic
/ APPLICANT: Wang, Youzhen
/ APPLICANT: Xu, Yongyao
/ APPLICANT: Hoersch, Sebastian
/ APPLICANT: Monahan, John
/ APPLICANT: Meyers, Rachel E.
/ APPLICANT: Bast Jr., Robert C.
/ APPLICANT: Hortobagyi, Gabriel N.
/ APPLICANT: Pusztai, Lajos
/ APPLICANT: Meric, Funda
/ APPLICANT: Sahin, Aysegul
/ APPLICANT: Mills, Gordon B.
/ TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
/ TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
/ FILE REFERENCE: MRI-038
/ CURRENT APPLICATION NUMBER: US/10/177,293
/ CURRENT FILING DATE: 2002-06-21
/ PRIOR APPLICATION NUMBER: US 60/299,887
/ PRIOR FILING DATE: 2001-06-21
/ PRIOR APPLICATION NUMBER: US 60/301,572
/ PRIOR FILING DATE: 2001-06-27
/ PRIOR APPLICATION NUMBER: US 60/306,501
/ PRIOR FILING DATE: 2001-07-18
/ PRIOR APPLICATION NUMBER: US 60/325,002
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: US 60/362,585
/ PRIOR FILING DATE: 2002-03-05
/ PRIOR APPLICATION NUMBER: US 60/xxx,xxx
/ PRIOR FILING DATE: 2002-05-14
/ NUMBER OF SEQ ID NOS: 506
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 11
/ LENGTH: 2872
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-177-293-211

Alignment Scores:
Pred. No.: 4,79e-20 Length: 2872
Score: 406.00 Matches: 142
Percent Similarity: 45.64% Conservative: 41
Best Local Similarity: 35.41% Mismatches: 146
Query Match: 19.12% Indels: 75
DB: 14 Gaps: 13

US-10-087-080-32 (1-402) x US-10-177-293-211 (1-2872)
Qy 24 GlyCysSerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43
Db 349 GGTCTGGCGGGCCCATGACACATGACATGCGCGCGGTGACGCGCCATGGGTACG--- 405
Qy 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaAargAspThrGlnGlyAspGly 63
Db 406 -----CGCTGAGCCCGAGC---GGCATGGCGCCATGGTCCGACGAGCGGGCC 453
Qy 64 GluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIleProAlaAlaAlaAla 83
Db 454 TCCATGATGAATGGCTTGGGCCCTTACGGCGCGCCCATGAACCCGTGATGAGCCCCATG 513
Qy 84 AlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyAlaGly 103
Db 514 GCGTACGGCGGCTCACTTGGCGCGCAGCGCGCGGC-----GGCGGCGCGCAG 564
Qy 104 SerGlyGluGlyAlaAargSerLysProTyThrArgProLysProProTyrSerTyr 123
Db 565 GCCAAGACGTTCAAGCGCAGTTACCGCAGCC-----AAGCGCCCTACTCGTAC 615

```

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 31, 2004, 05:50:04 ; Search time 2579 Seconds  
(without alignments)  
4654.749 Million cell updates/sec

Title: US-10-087-080-32

Perfect score: 2123

Sequence: 1 MKLEVFVPRAAHGDKQGSDDL.....AALVRRPGHLSYPVETLLA 402

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DBV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10087080/runat\_29032004\_114946\_24430/app\_query.fasta\_1.583  
-DB=EST -OPM=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -TRANS=human40.cdi -LIST=45  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -MINLEN=0 -MAXLEN=2000000000  
-DOALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087080.GCEN\_1.1.3549 @runat\_29032004\_114946\_24430 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vln:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	954	44.9	904	13	BQ922461
2	693	32.6	1117	14	CK028187
3	682	32.1	946	14	CD754754
4	676	31.8	799	14	CF997203
5	656.5	30.9	559	9	AI169632
6	615	29.0	804	14	CK030218
7	614	28.9	581	13	BQ285521
8	613.5	28.9	805	13	EX084216
9	611	28.8	564	13	BQ449254
10	586.5	27.6	514	10	BF282916
11	586	27.6	682	9	AL636071
12	566	26.7	878	29	CNS044M2
13	563.5	26.5	398	9	AI586081
14	560.5	26.4	688	12	BI443539
15	519	24.4	542	14	CD282719
16	507.5	23.9	518	10	BB637563
17	503.5	23.7	571	13	BQ615802
18	502	23.6	557	12	BM531626
19	497.5	23.4	590	12	BM573553
20	497.5	23.4	592	12	BM573447
21	497.5	23.4	596	12	BM586827
22	496	23.4	573	12	BM572519
23	493	23.2	425	10	BF290883
24	493	23.2	580	12	BM574529
25	493	23.2	582	12	BM530633
26	493	23.2	584	12	BM585032
27	490.5	23.1	556	12	BM573325
28	490.5	23.1	557	12	BM532671
29	490.5	23.1	564	12	BM532182
30	490.5	23.1	585	13	BQ480599
31	488.5	23.0	541	12	BM572533
32	488.5	23.0	560	12	BM574466
33	488.5	23.0	562	12	BM531746
34	487.5	23.0	571	12	BM532199
35	487.5	23.0	556	12	BM572314
36	487.5	23.0	570	12	BM574490
37	486.5	22.9	564	12	BM574918
38	486.5	22.9	570	12	BM532720
39	486.5	22.9	571	12	BM573745
40	485.5	22.9	565	12	BM572374
41	485.5	22.9	570	12	BM532670
42	484.5	22.8	554	12	BM574905
43	483.5	22.8	533	13	BQ615301
44	483.5	22.8	589	12	BM574575
45	483	22.8	561	12	BI709264

ALIGNMENTS

RESULT 1  
BQ922461  
LOCUS  
DEFINITION BQ922461 904 bp mRNA linear EST 20-AUG-2002  
IMAGE:6200329 5', mRNA sequence.  
ACCESSION BQ922461  
VERSION BQ922461.1 GI:22337492  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 904)

AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LAM13515 row: a column: 02  
 High quality sequence stop: 523.  
 Location/Qualifiers  
 1. 904  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6200329"  
 /sex="male"  
 /tissue\_type="sciatic nerve"  
 /dev\_stage="adult, 70 yr"  
 /lab\_host="DH10B"  
 /clone\_lib="Lupski sciatic nerve"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1: Noti; Site\_2: SalI; cDNA made by oligo-dr priming. Directionally cloned using the following adaptors: 5'-TCGACCGACGCGTCCG-3' and 5'-GACATGTTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.87 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

ORIGIN

Alignment Scores:  
 Pred. No.: 3.63e-32 Length: 904  
 Score: 954.00 Matches: 199  
 Percent Similarity: 70.10% Conservatives: 5  
 Best Local Similarity: 68.38% Mismatches: 29  
 Query Match: 44.94% Indels: 58  
 DB: 13 Gaps: 5

US-10-087-080-32 (1-402) x BQ922461 (1-904)

Qy 164 AnSerValArgHisAsnLeuSerLeuAsnAspCysPheVallyValLeuArgAspPro 183  
 |||||  
 Db 3 AACTCCGTGCGCCACACCTTTCGCTCAACGACTGCTTTGTCAAGGTGCTGCGCGACCCC 62  
 |||||  
 Qy 184 SerArgProTTPGLyLysAspAsnTyrTTPMetLeuAsnProAsnSerGluTyrThrPhe 203  
 |||||  
 Db 63 TCGCGGCCCTGGGCAAGGACACTACTGGATGCTCAACCCACACGAGTACACCTTC 122  
 |||||  
 Qy 204 AlaAspGlyValPheArgArgArgLysArgLysArgLysSerHisArgAlaProValProAla 223  
 |||||  
 Db 123 GCCGACGGGTCTTCGCGCGCGCGCGCGAAGCGCTCAGCCACCGCGCGCGGTCCCCGCG 182  
 |||||  
 Qy 224 ProGlyLeuArgProGluAlaProGlyLeuProAlaAlaProAlaProAlaProAla 243  
 |||||  
 Db 183 CCGCGGCTGCGCGCGCGAGGCGCCCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCC 242  
 |||||  
 Qy 244 AlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAlaSerProAla 263  
 |||||  
 Db 243 GCCCGCGCTCGCCCGCATCGCTCGCCCGCGCGCGCGAGGAGCGCGCGCGCGCGCGCG 302  
 |||||  
 Qy 264 GlyLysPheSerSerSerPheAlaIleAspSerIleLeuArgLysProPheArgSerArg 283  
 |||||  
 Db 303 GCGAAGTTCTCCAGTCTCGCATCGCATCGACAGATCTTCGCGAAGCGCTTCGCGAGCGCG 362  
 |||||  
 Qy 284 ArgLeuArgAspThrAlaProGlyThrThrLeuGlnTTPGlyAlaAlaProCysProPro 303  
 |||||

Db 363 CGCCTCAGGGACACGGCCCCCGGGAGCAGCGCTTCAGTGGGCGCGCGCCCTGCCCGCGC 422  
 Qy 304 LeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuProLeuCys 323  
 |||||  
 Db 423 CTGCCCCGGTTCCCGCGCTCTCTCCCGGGTGCCTCGAGGGCGCTCTGCGGCTCTGC 482  
 |||||  
 Qy 324 AlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProThr 343  
 |||||  
 Db 483 GCGTACGGCGCGGAGCGCGCGCTGGCGCGCCGCCGCCGCCGCCGCCGCCGCCGCCG 542  
 |||||  
 Qy 344 -----AlaProProLeuLeuAlaPro 351  
 |||||  
 Db 543 CCGCGCCCCCGGCTTCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 602  
 |||||  
 Qy 352 LeuProAlaAlaAlaProAlaLysProLeuArg-GlyProAlaAlaGlyAlaHisL 371  
 |||||  
 Db 603 TCTCCCG 659  
 |||||  
 Qy 371 euTyr----- 372  
 |||||  
 Db 660 TGTTCGCCCTCCCGCTATCCCGCGCTTTGTTCCTTTGTTCCTTGAATGCC 719  
 |||||  
 Qy 372 ----- 372  
 |||||  
 Db 720 GCGCGCGCGAGCTCGCATGGAGGAGCATAGGCGCGCGCGCGCGCGCGCGCGCGCG 779  
 |||||  
 Qy 373 --CysProLeuArgLeuProAlaAlaLeuGlnAlaLeuValArgArgProGly---P 391  
 |||||  
 Db 780 CGTGTCCCGCGCGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 830  
 |||||  
 Qy 391 roHisLeuSerTyrProValGluThrLeu 400  
 |||||  
 Db 831 TTCACCTGTCCCG 859  
 |||||

RESULT 2  
 CK028187 1117 bp mRNA linear EST 26-NOV-2003  
 LOCUS AGENCOURT 16624312 NIH\_ZGC\_7 Danio rerio cDNA clone IMAGE:7052472  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CK028187  
 VERSION CK028187.1 GI:38554111  
 KEYWORDS EST:  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.

REFERENCE  
 1 (bases 1 to 1117)  
 NIH-MGC <http://mgs.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Len Zon, Harvard  
 cDNA Library Preparation: Open Biosystems  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LAM14826 row: j column: 22  
 High quality sequence stop: 835.  
 Location/Qualifiers  
 1. 1117  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="IMAGE:7052472"  
 /tissue\_type="whole body"  
 /lab\_host="DH10B"

FEATURES  
 source

/clone lib="NIH ZGC 7"  
 /note="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV;  
 Bulk tissue was collected from a whole adult individual  
 from the Tubingen strain. 1st strand cDNA was primed with  
 a Not I - oligo(dT) primer, double-stranded cDNA was  
 cloned into the Not I and EcoRV sites of pExpress-1.  
 Library was size-selected for >1 kb fragments and  
 normalized. A non-normalized version of this library is  
 also available (NIH ZGC 10). Library was constructed by  
 Open Biosystems (Huntsville, AL)"

## ORIGIN

## Alignment Scores:

Pred. No.: 8,62e-21 Length: 1117  
 Score: 693.00 Matches: 167  
 Percent Similarity: 52.00% Conservative: 28  
 Best Local Similarity: 44.53% Mismatches: 78  
 Query Match: 32.64% Indels: 102  
 DB: 14 Gaps: 12

US-10-087-080-32 (1-402) x CK028187 (1-1117)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
 Db 64 ATGAACCTGGAGGTTTCTGCG-----GGGGTCACTACGACTCCAGCTCGGAGCTG 117  
 Qy 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40  
 Db 118 TGCAGCGATCTCAGGGAGCATCCATCCGCGGTGTCGCA-----GAGGAGGAGCTG 171  
 Qy 41 GlySerAspClyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 50  
 Db 172 GGCTCGATGGAGACTGCGTGGCGCACAGTCGCGCA-----CCTGTCGCCGACACCAA 225  
 Qy 61 GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluAlaAlaProAlaAla 80  
 Db 226 GGC----- 228  
 Qy 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100  
 Db 228 ----- 228  
 Qy 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120  
 Db 229 -----AAACCTACACTCGGAGACCCAACTCCA 258  
 Qy 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140  
 Db 259 TACTCTTACATCGACTTATCGCCATCGCCATCCGAGACTCCAACTCCGCGGACTCACT 318  
 Qy 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrThr 160  
 Db 319 CTAGCCGAATCAACGACTACCTCATGAGAGAGTTCCCGTTTTTTAGAGCGGACTACAC 378  
 Qy 161 GlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180  
 Db 379 GGCTGGAGGAACCTCAGTGGCCCAATATCTGCTCTAAAGACTGCTTTCTCAAGGCTT 438  
 Qy 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200  
 Db 439 AGGATCCTTCGAGACCGTGGGAAGGACAAATCTGATGCTGAACCGGACCGAG 498  
 Qy 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAlaPro 220  
 Db 499 TACACCTTCGGGACGGAGGTTCGACAGAGGAGAGAGCGCATTTAGTAAAAAACCC--- 555  
 Qy 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240  
 Db 556 -----GGCAGGGAGCCAGAG-----GGGCGGTGCA 582  
 Qy 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluAlaArgAla 260  
 Db 583 ACCACGCATGGACGATGATGATCCATCGCTACGCT----- 621

Qy 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280  
 Db 622 CTTTCGAGCGGAAGTTTCCAGAGTTCTTTTGGCATTGGAGGATCTCTCAGTCGACCTTC 681  
 Qy 281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro 300  
 Db 682 AGAAGGAGGAGACCCCGCTGCTCAGCCCTGACACC-----TGCCCGGGGGAGTG 732  
 Qy 301 CysProProLeuProAlaPheProAlaLeuProAlaAlaProCysArgAlaLeuLeu 320  
 Db 733 GACACTGTGTCGCG-----TATGTC 753  
 Qy 321 ProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAla----- 338  
 Db 754 ATGCTTCGATCCTATGTCGCTCGAGGACCTTCCATCGAGCGCCGACGCGGACTTT 813  
 Qy 339 -----GluValProThrAlaProProLeuLeuAlaProLeuProAlaAla 355  
 Db 814 TTCGATTCAGCTGCGCGCTGAGTGC-----CTCTCAATCCCGCATACGAGGCG 864  
 Qy 356 AlaProAlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHis 370  
 Db 865 GCGGTTGCC-----GCCCTGCAACCGCGGATATCAC 897  
 RESULT 3  
 CD754754 946 bp mRNA linear EST 30-JUN-2003  
 LOCUS AGENCOURT 14618637 NCI CGAP ZEMB2 Danio rerio cDNA clone  
 DEFINITION IMAGE:6964741 5', mRNA sequence.  
 ACCESSION CD754754  
 VERSION CD754754.1 GI:32339041  
 KEYWORDS EST.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 946)  
 NIH-MGC http://mgs.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Daniela S. Gerhard, Ph.D.  
 COMMENT Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Leonard I. Zon, M.D.  
 cDNA Library Preparation: Invitrogen Corp  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLM14605 row: c column: 12  
 High quality sequence start: 15  
 High quality sequence stop: 663.  
 Location/Qualifiers  
 1. 946  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="IMAGE:6964741"  
 /tissue\_type="embryo"  
 /lab\_host="DH10B (Ti-resistant)"  
 /clone\_lib="NCI CGAP ZEMB2"  
 /note="Vector: PCMV-SPOrt6.1; Site 1: EcoRV; Site 2: NotI;  
 Cloned unidirectionally. Primer: Oligo dT. Average insert  
 size 2 kb. Constructed by J. Wang (Research Genetics  
 Invitrogen Corp) from tissue donated by L. Zon (Harvard  
 University). Note: this is a NCI CGAP Library."

## FEATURES

## source

## ORIGIN

## Alignment Scores:

```

Pred. No.: 2.15e-20 Length: 946
Score: 682.00 Matches: 165
Percent Similarity: 52.59% Conservatives: 28
Best Local Similarity: 44.96% Mismatches: 75
Query Match: 32.12% Indels: 99
DB: 14 Gaps: 11

US-10-087-080-32 (1-402) x CD754754 (1-946)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db ATGAACCTGGAGGTTTCTGCG-----GGGGTCACCTACCGCTCCAGCCTCGGAGCTG 129

Qy 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40
Db TGCAGCGATGCTGAGGCGAGCATCCCATCGCGGTGTCGCA-----GAGGAGGAGCTG 183

Qy 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60
Db GGCTCGGATGGAGACTGGTGGCGCACAGTCCGCA-----CCTGTCGCCGACACCAA 237

Qy 61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaLeuProAlaAla 80
Db GGC----- 240

Qy 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaGlyProGlyAlaGly 100
Db ----- 240

Qy 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120
Db -----AAACCCCTACACTCGGAGACCCAAACCTCCA 270

Qy 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleAlaArgSerAlaGlyGlyArgLeuThr 140
Db TACTCTTACATCGCATTTATCGCATGCCATCGAGACTCCAGACTCCCGCGCGACTCACT 330

Qy 141 LeuAlaGluIleAenGlyTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160
Db CTAGCCGAATCAACGACTACCTCATGAGAGTTCCTCGTTTITAGAGGCAGCTACACC 390

Qy 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180
Db GGCTGGAGGAACTCAGTGGCCCAATACTCTCTAAACGACTGCTTTCTCAAGGCTTAA 450

Qy 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlu 200
Db AGGATCTCTCGACCGTGGGAAAGACAAATTACTGGATGCTGAACCGCGCACAGCGAG 510

Qy 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysSerHisArgAlaPro 220
Db TACACCTTCGGGACGGAGTGTTCGCGAGAGGAGAAAGCGCATAGTAAATAAACCC--- 567

Qy 221 ValProAlaProGlyLeuArgProGluAlaProGlyLeuProAlaAlaProProPro 240
Db -----GGCCGCGTGCRA 594

Qy 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAla 260
Db ACCCAGCATGGACAGCAATGACTCCATCGCTACGCT----- 633

Qy 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280
Db CTTTCGAGCGGANNAGTTCACAGTTCCTTCGCATTGAGAGCATCTTCAGTCAGCCCTTC 693

Qy 281 ArgSerArgArgLeuArgAsp-ThrAlaProGlyThrThrLeuGlnThrProGlyAlaAlaPr 300
Db AGA-----AGGAGGACCGCCCGTGTCTCAGCCCTGACACCTCGCCGGGGGA 741

Qy 300 o-----CysProProLeuProAlaPheProAlaLeuLeuProAlaAla----- 314
Db GTGGACACTGTGCGCGCGGTATGTCTATCGATCTTATGTCCTTATGTCCTTGNAGACCCNNNC 801

```

```

Qy 315 -----ProCysArgAlaLeuLeuProLe 322
Db 802 NNGAGNCGCGANGCGCGACTTTTTCGCATTTCAGTCGCGCTGGGATGCTCTCATCCNG 861
Qy 322 uCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProPr 342
Db 862 NTACNAGCGCGCTGCGCGCACCTTCGACGCGGGATTCCACGTCANTTGTATCTCTGCC 921

Qy 342 oThrAlaProProLeuLeu 348
Db 922 NACN-----CGGTTACTC 934

```

## RESULT 4

CF997203

LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION

CF997203

VERSION

CF997203.1

KEYWORDS

EST

SOURCE

ORGANISM

Danio rerio (zebrafish)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabbs-remail.nih.gov

Tissue Procurement: Len Zon, Harvard

cDNA Library Preparation: Open Biosystems

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Place: LLM14784 row: C column: 19

High quality sequence start: 10

High quality sequence stop: 715

Location/Qualifiers

1..799

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clones="IMAGE:7036173"

/tissue\_type="whole body"

/lab\_host="DH10B"

/clone\_lib="NIH\_ZGC\_7"

/note="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV;

Bulk tissue was collected from a whole adult individual

from the Tuebingen strain. 1st strand cDNA was primed with

a Not I - oligo(dT) primer, double-stranded cDNA was

cloned into the Not I and EcoRV sites of pExpress-1.

Library was size-selected for &gt;1 kb fragments and

normalized. A non-normalized version of this library is

also available (NIH\_ZGC\_10). Library was constructed by

Open Biosystems (Huntsville, AL)"

ORIGIN

Alignment Scores:

Pred. No.: 3.24e-20

Score: 676.00

Percent Similarity: 54.88%

Best Local Similarity: 47.87%

Query Match: 31.84%

DB: 14

Length: 799

Matches: 157

Conservative: 23

Mismatches: 52

Indels: 96

Gaps: 11



337 GluAlaGluValProThrAlaProProLeuLeuAlaProLeuProAlaAlaAla 356  
 137 GGGCCGAGGTGAACCGCGGG---CCCTGTGCTGGCGCCCTCCACCGGGCC 81  
 357 ProAlaAlaProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeu 375  
 80 CCAGCAAGCCATTTCAGGTCCGGAGACCGCGCGCGCGCACCTGTACTGCCCCCTA 21  
 376 ArgLeuProAlaAlaLeu 381  
 20 CGGCTGCCACCGCCCTG 3

RESULT 6  
 CK030218 804 bp mRNA linear EST 26-NOV-2003  
 LOCUS AGENCOURT\_16624545 NIH\_ZGC\_7 Danio rerio cDNA clone IMAGE:7051925  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CK030218  
 VERSION CK030218.1 GI:38556142  
 KEYWORDS EST.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 804)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Len Zon, Harvard  
 CDNA Library Preparation: Open Biosystems  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Date: LAM14825 row: d column: 03  
 High quality sequence stop: 716.

FEATURES  
 source  
 1..804  
 Location/Qualifiers  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7965"  
 /clone="IMAGE:7051925"  
 /tissue\_type="whole body"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_ZGC\_7"  
 /notes="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV;  
 Bulk tissue was collected from a whole adult individual  
 from the Tuebingen strain. 1st strand cDNA was primed with  
 a Not I - oligo(dT) primer, double-stranded cDNA was  
 cloned into the Not I and EcoRV sites of pExpress-1.  
 Library was size-selected for >1 kb fragments and  
 normalized. A non-normalized version of this library is  
 also available (NIH\_ZGC\_10). Library was constructed by  
 Open Biosystems (Huntsville, AL)"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1..41e-17 Length: 804  
 Score: 615.00 Matches: 141  
 Percent Similarity: 58.36% Conservative: 16  
 Best Local Similarity: 52.42% Mismatches: 40  
 Query Match: 28.97% Indels: 72  
 DB: 14 Gaps: 8

US-10-087-080-32 (1-402) x CK030218 (1-804)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
 Db 95 ATGAAGTTGGAGGTTTC-----TCTCGGAGTCCCTTGTGGACAGCCCTCGACCTG 148  
 QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40  
 Db 149 TGCAGTGACATGGATGTGAACGTGCTGCCTTCCCGTGTCCACG-----GAGGAGGAATTG 202  
 QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60  
 Db 203 GGCTCGGATGGAGATGCTCAGCCAAC----- 229  
 QY 61 GlyAspGlyGluGlnSerAlaAlaGlyGlyProGlyAlaGluGluAlaAlaProAlaAla 80  
 Db 230 -----AGTCCAGGACCGGTGCCCT----- 250  
 QY 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100  
 Db 251 -----GTTCCGGACGG----- 262  
 QY 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120  
 Db 263 -----AGGCGAAACCATATACGCGTAGACCCCAACCACT 298  
 QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140  
 Db 299 TACTCTTACATTGCACCTATTGCAATGGCCATACGCGATTCCAAACACCGGTGCTTTACA 358  
 QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrThr 160  
 Db 359 CTTCCTGGAATAAAGCAATCTTAATGAAAAATCCCGTTTTTCGGGCGAGCTACACG 418  
 QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180  
 Db 419 GGGTGGCGAAACTCTGTACGCATAAATTGTCTATAAACCACTGCTTTTAAAAAGTTTGG 478  
 QY 181 ArgAspProSerArgProTropGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200  
 Db 479 CGAGATCCCTCGCTCGGTGGGCAAGACAAATTTTGGATGTTGAACCCACACAGCGGAG 538  
 QY 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysLeuSerHisArgAlaPro 220  
 Db 539 TACACTTTTGCAGACGGCGTTTCCCGCGGAGGAGGAAGCGCATCAGT----- 586  
 QY 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240  
 Db 587 -----AGAGATCTCGGCGCGCGGGAATCCCCAGAGCGC 622  
 QY 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluAlaAla 260  
 Db 623 GTGCCCGCGGATGACAGC-----AGGCTTCCAGCGCGGCGAGAAAGC----- 664  
 QY 261 SerProAlaGlyLysPheSerSerSer 269  
 Db 665 -----GTCTCTAAAGTTTTCAGACTCT 685

RESULT 7  
 BQ285521 581 bp mRNA linear EST 14-MAY-2002  
 faas1809.y1 zebrafish fin day3 regeneration Danio rerio cDNA clone  
 IMAGE:5942369 5', similar to SW:HFHL\_RAT\_Q83244 HEPATOCYTE NUCLEAR  
 FACTOR 3 FORKHEAD HOMOLOG 1 ; mRNA sequence.  
 LOCUS BQ285521  
 DEFINITION BQ285521.1 GI:20654021  
 VERSION BQ285521  
 KEYWORDS EST.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 581)  
 Clark M., Johnson, S.L., Leirach, H., Lee, R., Li, F., Marra, M.,  
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R., and Wilson, R.  
 WashU Zebrafish EST Project 1998  
 Unpublished (1998)  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrfish@watson.wustl.edu  
 CDNA Library Preparation: Raymond Lee. CDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
 Missouri (web address: www.genomesystems.com) (email contact:  
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
 (web address: www.resgen.com) (email contact: info@resgen.com) and  
 ResourcenetrumPrimarDatenbank, Berlin, Germany (web address:  
 www.rzpd.de)  
 Seq primer: T3 ET from Amersham  
 High quality sequence stop: 487.

# FEATURES

Location/Qualifiers

1..581

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="IMAGE:5912369"

/sex="mixed male and female"

/issue\_type="3 day fin regenerates"

/lab\_host="E. coli XL0LR"

/clone\_lib="zebrafish fin day3 regeneration"

/note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; 1st  
 strand cDNA primed with (GA)10ACTAGTCGAG(T)18, followed  
 by second strand synthesis, and ligated to 5' adapter  
 (5')-aattgcgcagcg-3', 3'-gccgtgtc-5'. cDNA was cloned  
 directionally (EcoRI/XhoI) into Stratagene Zap express  
 lambda phage arms. Mass invivo excision done to obtain  
 inserts in pBK-CMV phagemid."

## ORIGIN

### Alignment Scores:

Pred. No.: 1..09e-17 Length: 581  
 Score: 614.00 Matches: 129  
 Percent Similarity: 64.41% Conservative: 14  
 Best Local Similarity: 58.11% Mismatches: 26  
 Query Match: 28.92% Indels: 53  
 DB: 13 Gaps: 5

US-10-087-080-32 (1-402) x BQ285521 (1-581)

QY 109 ArgSerLysProTyrThrArgProLysProTyrSerTyrIleAlaLeuIleAla 128  
 Db 35 AAAGGCAACCTTACCTCGGAGACCAACCTCCATCTTATCATCGCATTCGCC 94  
 QY 129 MetAlaIleArgAspSerLaglyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeu 148  
 Db 95 ATGGCCATCCGAGACTCCAACTCCGCGCAGCTCACTTAGCGGAAATCAAGACTACCTC 154  
 QY 149 MetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSerValArgHis 168  
 Db 155 ATGAAGAAGTCCCGTTTTTATAGACAGCTACACCGCTGGAGAACTCATGTGGCCAT 214  
 QY 169 AsnLeuSerLeuAsnAspCysPheValLysValLeuAtgAspProSerArgProTyrGly 188  
 Db 215 AATCTGCTCTAAACGACTGCTTCTCAGGCTCTTAAGGGATCCTTCGAGACCGGGGGA 274  
 QY 189 LysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPhe 208  
 Db 275 AAGGACAATCTAGTGTGTGACCCCGCAGCAGTACACCTTCGCGACGGAGTGT 334  
 QY 209 ArgGArgArgLysArgLeuSerHisArgAlaProValProAlaProGlyLeuArgPro 228

Db 335 CGCAGAGGAGAAAGCGCATTAGTAAAAAAC-----GGCAGGGAGCCA 379  
 QY 229 GllclulalProGlyLeuProAlaAlaProProAlaProAlaAlaProAlaSerPro 248  
 Db 380 GAG-----GGCCCGTGCACCAACCCAGCATTTGGACAGCAATGAC 418  
 QY 249 ArgMetArgSerProAlaArgGlnGluArgAlaSerProAlaGlyLysPheSerSer 268  
 Db 419 TCCATCGTACGCT-----CCTTCGAGCGGAAAGTTCCACCACT 457  
 QY 269 SerPheAlaIleAspSerIleLeuArgLysProPheArg-SerArgLeuArgAspTh 288  
 Db 458 TCCTTCGCCATTGAGAGCATCCTCAGTCGACCTTCAGAGGGAGG----- 503  
 QY 288 rAlaProGlyThrThrLeuGlnTrpGlyAlaAlaProCysProProLeuProAlaPhePr 308  
 Db 503 ----- 503  
 QY 308 oAlaLeuProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGl 328  
 Db 504 -----ACCGCCCGTGTCTCAGCCCTGACACCT-----GCCCGGG 538  
 QY 328 YGlu 329  
 Db 539 GGAG 542

### RESULT 8

EX084216

LOCUS

DEFINITION

EX084216 805 bp mRNA linear EST 25-APR-2003

EX084216 AGENAE Rainbow trout normalized multi-tissues library

(tcac) Oncorhynchus mykiss cDNA clone tcac0005c.1.17 5prim, mRNA

sequence.

EX084216

KEYWORDS

SOURCE

ORGANISM

Oncorhynchus mykiss (rainbow trout)

Oncorhynchus mykiss

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 805)

Govoroun, M., Guiguen, Y. and Le Gac, F.

Construction and primary characterization of normalized cDNA

libraries in rainbow trout, Oncorhynchus mykiss

Unpublished (2003)

Contact: Guiguen Y

INRA - SCRIBE

Campus de Beaulieu, RENNES cedex, 35042, France

Tel: 02.23.48.50.09

Fax: 02.23.48.50.20

Email: Yann.Guiguen@beaulieu.rennes.inra.fr

Sequence cleaned of vector, adaptor and repetitions. Contact us

at signasupport@jouy.inra.fr to obtain the chromatogram of this

sequence.

Plate: 0005 row: 1 column: 17

Seq primer: T7.

Location/Qualifiers

1..805

/organism="Oncorhynchus mykiss"

/mol\_type="mRNA"

/db\_xref="taxon:8022"

/clone="tcac0005c.1.17"

/tissue\_type="adipose tissue, blood, brain,

differentiating gonads, gills, interrenal, intestine,

kidney, liver, muscle, ovary, pituitary, testis"

/dev\_stage="from embryos to adults"

/lab\_host="DH10B"

/clone\_lib="AGENAE Rainbow trout normalized multi-tissues

library (tcac)"

/notes="Vector: pT73D-pac; Clone distribution : AGENAE

Resource Centre, Francois PIUMI,

Francois.PIUMI@jouy.inra.fr, INRA, CEA Radiobiologie et

Etude du genome (LREG), Domaine de Vilvert, 78352,

ORIGIN  
Jouy-en-Josas cedex, France"

Alignment Scores:  
Pred. No.: 1,428-17 Length: 564  
Score: 611.00 Matches: 128  
Percent Similarity: 61.14% Conservative: 12  
Best Local Similarity: 55.90% Mismatches: 29  
Query Match: 28.90% Indels: 60  
DB: 13 Gaps: 5

US-10-087-080-32 (1-402) x BQ49254 (1-805)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGly----- 17  
Db 174 ATGAACCTTGAATTTCTGTGGAAC---CACCAACAGGAGTGAATCTTTCAGGAG 230

Qy 18 -----SerAspLeuGluGlyAlaGlySerAspAlaProSerProLeuSerAlaAla 35  
Db 231 CTGTGACGCGATCCGAGGGGAGC-----GTCCGCTGCGCTCTCTCTCC----- 272

Qy 36 GlyAspSerLeuGlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGly 55  
Db 273 GGTGAGGAGAGTGGTTCGATGGGACTGCTGCGG----- 311

Qy 56 AlaArgAspThrGlnGlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGlu 75  
Db 312 -----CAC 314

Qy 76 AlaIleProAlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAla 95  
Db 315 AGCCGCCACCTGTCACA----- 332

Qy 96 GlyProGlyAlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArg 115  
Db 333 ---CCCGCGTAGACGGCATAGGATTGGC-----AAACCGTATATACGG 374

Qy 116 ArgProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAla 135  
Db 375 AGACCAAACTCCGTACTCTTACATGCATGATTCGCTATCCGATCCGCTGACTCAACA 434

Qy 136 GlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePhe 155  
Db 435 TCTGTCGTCTGACTCTGGCAGAGATCAATGACTATCTGATGAAAAGTTTCCATTTC 494

Qy 156 ArgGlySerTyrThrGlyTyrArgAsnSerValArgHisLeuSerLeuAsnAspCys 175  
Db 495 CGGGCAGCTACACTGGTTGAGAAATTCGTCGCGCCACACTTGTCACTTAACGACTGT 554

Qy 176 PheValLysValLeuAlaArgProSerArgProTyrGlyLysAspAsnTyrTyrMetLeu 195  
Db 555 TTCTTAAGTCTCTCCGGGACCCGTCAGACCCCTGGGAAAAGACAACACTACTGATGCTG 614

Qy 196 AsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeu 215  
Db 615 AATCTCAACAGAGTACACCTTCGTCACGAGGTGTTCGACGAGGAGAAACGCGATC 674

Qy 216 SerHisAlaProValProAlaProGlyLeuArgPro-----GluGluAla 231  
Db 675 AATAAAAAGACAGG---CAAGAGCAAAACGCTCCGACCATCTCCGACGAGACGGCA 733

Qy 232 Pro-GlyLeuProAlaAlaProProAlaProAlaProAlaProAlaSerProArgMetAr 251  
Db 734 AGTCACCATCATCATCCACCCCTCCACCACTAAACCGNACCCAGCAGT----- 782

Qy 251 gSerProAlaArgGlnGluGluArgAlaSerProAlaGlyLysPheSerSerPhe 270  
Db 783 -----GCCAAGTCTCCAGTTCCTTC 803

RESULT 9  
BQ49254  
LOCUS faa45a03.y1 zebrafish fin day3 regeneration Danio rerio cDNA clone  
DEFINITION

IMAGE:5911949 5' similar to TR:070220 O70220 FORK HEAD  
TRANSCRIPTION FACTOR. ; mRNA sequence.

ACCESSION BQ49254  
VERSION BQ49254.1 GI:21252366  
EST: Danio rerio (zebrafish)  
SOURCE Danio rerio  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 564)  
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,  
Edg, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.  
WashU Zebrafish EST Project 1998  
Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu  
cDNA Library Preparation: Raymond Lee. cDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:  
www.rzpd.de)  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 450.  
Location/Qualifiers  
1..564  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:5911949"  
/sex="mixed male and female"  
/tissue\_type="3 day fin regenerates"  
/lab\_host="E. coli XL0LR"  
/clone\_lib="zebrafish fin day3 regeneration"  
/note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; 1st  
strand cDNA primed with (GA)10ACTAGTCTCGAG(T)18, followed  
by second strand synthesis, and ligated to 5' adapter  
(5'-aactggcagag-3', 3'-gccgtgctc-5'. cDNA was cloned  
directionally (EcoRI/XhoI) into Stratagene Zap express  
lambda phage arms. Mass invivo excision done to obtain  
inserts in pBK-CMV phagemid."

ORIGIN  
Alignment Scores:  
Pred. No.: 1,428-17 Length: 564  
Score: 611.00 Matches: 128  
Percent Similarity: 61.14% Conservative: 12  
Best Local Similarity: 55.90% Mismatches: 29  
Query Match: 28.90% Indels: 60  
DB: 13 Gaps: 5

US-10-087-080-32 (1-402) x BQ49254 (1-564)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
Db 9 ATGAACCTGAGGTTTCTGCG-----GGGGTCTACTAGACTCCAGCCTCGGAGCTG 62

Qy 21 GluGlyAlaGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40  
Db 63 TGACGCGATGCTGAGGGGAGCATCCATCCCGGTGTCGCA-----GAGGAGGCTG 116

Qy 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60

```

Db      117 GCCTGGATGAGACTGGTGGCGCACAGTCGGCA-----CCTGTCCGCACACCAA 170
Qy      61 GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGluAlaProAlaAla 80
Db      171 GGC----- 173
Qy      81 AlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100
Db      173 ----- 173
Qy      101 GlyAlaGlySerGlyGluGlyAlaArgSerIysProTyrThrArgArgProGlyPro 120
Db      174 -----AAACCTACACTCGAGACCCAAACCTCCA 203
Qy      121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyValArgLeuThr 140
Db      204 TACTTTTACATCGCACTTATCGCCATGCGCATCCAGACTCCAACTCCGGCGCACTCACT 263
Qy      141 LeuAlaGluIleAsnGluTyrIleuMetGlyLysPheProPheArgGlySerTyrThr 160
Db      264 CTAGCCGAATCAACGCACTACCTCATGAAGAAGTTCCTGTTTATAGAGGAGCTACACC 323
Qy      161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeu 180
Db      324 GGCCTGGAGAACTCAGTGGCGCATTAATCTGTCTCTAAACGCACTGCTTTCTCARGGTCTTA 383
Qy      181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlu 200
Db      384 AGGGATCCTTCGAGACCGTGGGAAAGGACAACTACTGATGCTGATGATGATGATGATGAT 443
Qy      201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysArgLysArgLysArgLys 220
Db      444 TACACCTTCGCGAGCGAGTGTTCGCAGAGGAGAAAGCGCATTAGTAAAGAAACCC 500
Qy      221 ValProAlaProGlyLeuArgProGlu 229
Db      501 -----GGCAGGAGCCAGAG 515

```

## RESULT 10

BF282916/c

LOCUS

DEFINITION EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

```

/lab host="DH5-alpha"
/clone lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/notes="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RLI, RPL, RLU, REM, RMD,
RSP, RHE, RPC, RPN"

ORIGIN
Alignment Scores:
Pred. No.: 1.47e-16 Length: 514
Score: 586.50 Matches: 125
Percent Similarity: 75.58% Conservative: 5
Best Local Similarity: 72.67% Mismatches: 39
Query Match: 27.63% Indels: 3
DB: 10 Gaps: 3

US-10-087-080-32 (1-402) x BF282916 (1-514)

Qy      213 LysArgLeuSerHisArgAlaProValProAlaProGlyLeuArgProGluGluAlaPro 232
Db      513 AAGCGGCTCAGCCACCGGACCAAGTCTCCGATCCGGGCTACGGCCGAGGAGGCCCA 454
Qy      233 GlyLeuProAla--AlaProProAlaProAlaProAlaProAlaSerProArgMetArg 251
Db      453 CCGGACCTCGGGGACCCCGAGCCGCCGCCACCCCGGGCTCTCCCAATCCGCGCG 394
Qy      252 SerProAlaArgGlnGluArgAlaSerProAlaGlyLysPheSerSerPheAla 271
Db      393 TCGCCCGCTCGCCAGGAGGAGGCTCCAGCCGCGGAGCAAGTCTCCAGCTCTTCGCG 334
Qy      272 IleAspSerIleLeuArgLysProPheArgSerArgArgLeuArgAspThrAlaProGly 291
Db      333 ATCGACAGCATCTCTCAGCAAGCGGTTTCGAGCCGCCGCGGCGGCGGCGGCGG 274
Qy      292 ThrThrLeuGlnTyrGlyAlaAlaProCysProProLeuProAlaPheProAlaLeu 311
Db      273 GTCAGCTACCTGGAGCGCTGCTCTCCCTGCGCGCTGCGCGCTATCCCGGCTCTT 214
Qy      312 ProAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAla 331
Db      213 CCGCGCTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 154
Qy      332 ArgLeuGlyAlaArgGluAlaGluValProProThrAlaProProLeuLeuAlaPro 351
Db      153 CTCTGCGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 97
Qy      352 LeuProAlaAlaAlaProAlaLysProLeuArgGlyPro---AlaAlaGlyAlaHis 370
Db      96 CTCTCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 37
Qy      371 LeuTyrCysProLeuArgLeuProAlaAlaLeuGln 382
Db      36 CTGTACTGCCCCCTACGGCTGCCACGCGGCGCTGCAG 1

RESULT 11
AL636071
LOCUS
DEFINITION AL636071 XGC-neurula Silurana tropicalis cDNA clone TNeu015112 5',
mRNA sequence.
ACCESSION AL636071
VERSION AL636071.2 GI:38214106
KEYWORDS EST.
SOURCE Silurana tropicalis (western clawed frog)
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 682)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
On Nov 7, 2001 this sequence version replaced gi:16788050.
Contact: Huckle E

```

## FEATURES

source

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/db\_xref="taxon:10116"

/clone="RGIDU92"

/tissue\_type="mixed tissue"

Sanger Institute  
 Hinxton, Cambridgeshire, CB10 1SA, UK  
 Email: tropesanger.ac.uk  
 This sequence is from a Xenopus Gene Collection (XGC) library  
 constructed by Aaron M. Zorn.  
 cDNA was oligo dT primed from Sug of poly A+ RNA from neurula.  
 EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the  
 5' end and NotI at the 3' end.  
 Vector: pCS107; Site1: EcoRI; Site2: NotI  
 Host: Escherichia coli DH10B  
 Sanger Xenopus tropicalis EST project 2001  
 TROPICALIS SEQUENCE ID: TNeu015112.plkSP6  
 Sequencing primer: SP6

FEATURES  
 source  
 1..682  
 /organism="Silurana tropicalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8364"  
 /clone="TNeu015112"  
 /dev\_stage="neurula"  
 /lab\_host="Escherichia coli DH10B"  
 /clone\_lib="XGC-neurula"  
 /notes="Vector: pCS107; Site1: EcoRI; Site2: NotI; cDNA  
 was oligo dT primed from Sug of poly A+ RNA from neurula.  
 EcoRI-NotI cut cDNA was then ligated into pCS107 with  
 EcoRI at the 5' end and NotI at the 3' end."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,11e-16 Length: 682  
 Score: 586.00 Matches: 129  
 Percent Similarity: 61.51% Conservativeness: 18  
 Best Local Similarity: 53.97% Mismatches: 50  
 Query Match: 27.60% Indels: 42  
 DB: 9 Gaps: 6

US-10-087-080-32 (1-402) x AL636071 (1-682)

Qy 62 AspGlyGluInSerAlaGlyGlyProGlyAlaGluGluAlaLeuProAlaAla 81  
 Db 7 GACACTGACACCGGTAGCTGGGGGGATGAAGAAGTAA----- 48  
 Qy 82 AlaAlaValValAlaGluGlyAlaGluAlaGlyProGlyAlaGlyGly 101  
 Db 49 -----GTGGAAGAAGAGGAGGAGTAAACCCAGAGAGAATGGGTCTCGGCA 96  
 Qy 102 AlaGlySerGly-----GluGlyAlaArgSerLysProTyrThr 114  
 Db 97 GATGGTTCTACGCAATGCCGAGCACAAATTTGTTAGGGGGCAACAAAGACTTACAC 156  
 Qy 115 ArgArgProLysProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSer 134  
 Db 157 CGCCTTTCGAACCTCCTTACTCCTACATCGCACTCATGCCATGCTATTAAAGACTCA 216  
 Qy 135 AlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhe 154  
 Db 217 CCAGTGGGGCGCTGACCCCTTCAGAGATCAATGACTATCTGATGAAGATTCCCATTC 276  
 Qy 155 PheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisnLeuSerLeuAsn 174  
 Db 277 TTCAGGGGTAGTACACGGCTGGAGAACTCGGTGCGACACAATCTTTCCTCCACAGAC 336  
 Qy 175 CysPheValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMet 194  
 Db 337 TGTTTCGTCAAGTTCTGGCGGACCCCATCAGCCCTTGGGGCAACACCACTACTGGATG 396  
 Qy 195 LeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArg 214  
 Db 397 CTGAACCCGAATAGTAGTACCTTCGACAGCGGCTTTCAGACGCAAGAGGAGCGCA 456  
 Qy 215 LeuSerHisArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeu 234  
 Db 457 CTTAACAGGGTCAACCAAGTGC-----CTCAAGAGCAGGATTTGCAGAGGCTT 504

Qy 235 Pro-----AlaAlaProPro 239  
 Db 505 GCAGACAGCAGCATCAATAATGATGAACCCCAACAAAGCCAGGATCTTCTCCCTCA 564  
 Qy 240 Pro-----AlaProAlaProAlaSerProArgMetArgSerProAla 254  
 Db 565 TCGAGCGGCTGATCATGGCACCTCAGCACCATCTCTCTACAAACTCTCTCTCC 624  
 Qy 255 ---ArgGlnGluGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIle 272  
 Db 625 AACAGGTCTGCCAAAGAAACCACTCANGGACCAAGTTTCCAGCTCTCTTGCATT 681

RESULT 12  
 CNS044M2  
 LOCUS  
 DEFINITION  
 Tetraodon nigroviridis genome survey sequence T7 end of clone  
 081F22 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 ACCESSION  
 AL274259  
 VERSION  
 AL274259.1 GI:7996530  
 KEYWORDS  
 GSS; genome survey sequence.  
 SOURCE  
 Tetraodon nigroviridis  
 ORGANISM  
 Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE  
 1 Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
 Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,  
 Saurin, W., and Weissenbach, J.  
 Estimate of human gene number provided by genome-wide analysis  
 using Tetraodon nigroviridis DNA sequence  
 Nat. Genet. 25 (2), 235-238 (2000)  
 20296633  
 PUBMED  
 10835845

TITLE  
 Estimate of human gene number provided by genome-wide analysis  
 using Tetraodon nigroviridis DNA sequence

JOURNAL  
 MEDLINE  
 PUBMED  
 20296633  
 PUBMED  
 10835845

REFERENCE  
 3 (bases 1 to 878)  
 Direct Submission  
 Genoscope.  
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 - Web : www.genoscope.cns.fr)  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES  
 source  
 1..878  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone="081F22"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : COBG081DC11LP1-end : T7"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,03e-15 Length: 878  
 Score: 566.00 Matches: 144  
 Percent Similarity: 52.48% Conservativeness: 15  
 Best Local Similarity: 47.52% Mismatches: 77  
 Query Match: 26.66% Indels: 68  
 DB: 29 Gaps: 9

```

US-10-087-080-32 (1-402) x CNS044M2 (1-878)
QY 30 SerProLeuSerAlaAlaGlyAspAspSerLeuGlySerAspGly----- 44
Db 95 TCGCCAAACAGCGCG-----TCGTGGGTTTCATGTGTCTGACGCGCAGAGA 142
QY 45 -----AspCysAlaAlaAlaPro 50
Db 143 CTTTITAGATGCCATATTGTGGCGCTGCTGCCGTAGAGGAGCATCTGCTATTAAACAT 202
QY 51 SerAlaGlyGlyAlaAlaG-----AspThrGlnGlyAspGlyGluGln 65
Db 203 GAAGCTGAGGTGTGTCTGCGCGCGCCACTATGATAAAAAACACGAGGTGACCGGAGGC 262
QY 66 SerAlaGlyGlyPro---GlyAlaGluGluAlaIleProAlaAlaAlaAlaAla 84
Db 263 GGAGGAGGAGCCT-CCCTGTCTGTTGAAGAGCGCTGGCGTCASTAGGAGCTTCGTG 321
QY 85 ValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyGlySer 104
Db 322 GTGCGCASCAGCAGCTGTGACTCCATCGCGC----- 354
QY 105 GlyGluGlyAlaAspSerIleProThrArgArgProLysProProTy-SerTyrIle 124
Db 355 ---GTGGACGCCCAAGTCTAAACCGTACATACGAGACCTAACCTCTTCTCTACATC 411
QY 125 -AlaLeuIleAlaMetAlaIleAlaArgAspSerAlaGlyArg-LeuThrLeuAlaGluI 144
Db 412 AGCCCTCATCGCATGTCGCAATTCGGGACTTCGGTACTGGCAGCGCTGATTCGGCGGAGA 471
QY 144 LeuGluThrLeuMetGlyIlePheProPheArgGlySerTyrThrGlyTrpArgA 164
Db 472 TCAACAACATCTGATGTCAGAAATTTCCATTTTGTAGAGGAGCGCACCGGCGTGGAGGA 531
QY 164 snSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeuArgAspPro 184
Db 532 ACTCTGTTGACACAACTGCTCTTAACGACTGCTTCTCAAGTGCTCCGCGACCGGT 591
QY 184 eArgProTrpGlyIleAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheA 204
Db 592 CCAGACCTTGGGAAAGACAAATTAAGTGTCTGAATCCCAACAGTACAGTATACCTTTK 651
QY 204 laAspGlyValPheArgArgArgGlyAlaGlyLeuSerHisArgAlaProValProAla 224
Db 652 CGATGGAGTGTTCGACCGCAGAGAAAGCGCATTCACAAAAA----- 695
QY 224 roGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaA 244
Db 696 ---CTAAACACGGAG-----CCAAACGGCGCGAG 720
QY 244 laProAlaSerProArgMetArgSerProAlaArgGlnGluGluAlaSerProAlaG 264
Db 721 AGCAGCAGACGCGCGCATTTTCCGCTGACCGACGACCGAT-----TCGTGCG 771
QY 264 lYlePheSerSerPheAlaIleAspSerIleLeuArgIlePheArgSerArgA 284
Db 772 GSAAGTWTAGAGTAGTATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 824
QY 284 rgLeuArgAspThrAlaProGlyThrLeuGluThrTrpGlyAlaAlaProCysProPro 304
Db 825 -----AGGAGCCCGCAGACACCATTTGTGAGCCCGCTCCTGCTCACTGCGCA 873
QY 304 euPro 305
Db 874 GCGCC 878

RESULT 13
LOCUS
DEFINITION
AI586081
IMAGE:1209608 3, similar to gb.U36760 Mus musculus brain factor-1
(MOUSE);, mRNA sequence.

```

ACCESSION AI586081  
 VERSION AI586081.1 GI:4571978  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 REFERENCE 1 (bases 1 to 398)  
 AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.  
 TITLE The WashU-NCI Mouse EST Project 1999  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:645952  
 This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
 High quality sequence stop: 387.  
 Location/Qualifiers  
 1..398  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1209608"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="Stratagene mouse skin (#937313)"  
 /notes="Organ: skin; Vector: pBluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dr. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5' CTCAGTCTTTTCTTTTCTTTT 3'"

Alignment Scores:  
 Pred. No.: 1..15 Length: 398  
 Score: 563.50 Matches: 111  
 Percent Similarity: 87.02% Conservative: 3  
 Best Local Similarity: 84.73% Mismatches: 16  
 Query Match: 26.54% Indels: 1  
 DB: Gaps: 1  
 US-10-087-080-32 (1-402) x AI586081 (1-398)  
 QY 162 TrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeuArg 181  
 Db 2 TGGCGCACTCCGTGGCGCACACCTCTCGCTCAACGACTGTTCTCAAGGTGCTGCGC 61  
 QY 182 AspProSerArgProTrpGlyIleAspAsnTyrTrpMetLeuAsnProAsnSerGluTyr 201  
 Db 62 GACCCCTCGCGCGCCTGGGCGCAAGGCAACTACTGTGATGCTCAACCCCAACGCGAATAC 121  
 QY 202 ThrPheAlaAspGlyValPheArgArgArgGlyAlaGlySerHisArgAlaProVal 221  
 Db 122 ACTTTCGCCGACGGGTCTTCGCCCGCGCGCAAGCGCTCAGCCACCGGACCAGATC 181  
 QY 222 ProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla---AlaProProPro 240  
 Db 182 TCGCGGTCCGGGTGCGGCGCGGAGGAAGGCCACCGGACCTCGCGGAGCCCCCGGAGCCCC 241

```

Qy 241 AlaProAlaAlaProAlaSerProAlaMetArgSerProAlaArgGlnGluGluAla 260
Db 242 GCGCCGCGCGCGCTCTCCCGATCGCGGCTCGCGGCTCGCGAGGAGCGCTCG 301
Qy 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProphe 280
Db 302 AGCCCTCGAGCAAGTCTCCAGCTCTCGCATCGACAGCATCTCAGCAAGCCTTTT 361
Qy 281 ArgSerArgLeuArgAspThrAlaProGly 291
Db 362 CGCAGCGCGCGAGCGGACTCGGCTCTGGG 394

RESULT 14
BI443539/c
LOCUS
DEFINITION
dag55b10.x1 Wellcome CRC PRN3 St19 26 egg animal cap Xenopus laevis
cDNA clone IMAGE:4785354 3', similar to SW.HF1 RAT Q63244
HPATOCYTE NUCLEAR FACTOR 3 FORKHEAD HOMOLOG 1-, mRNA sequence.
ACCESSION
BI443539
VERSION
BI443539.1 GI:15268246
KEYWORDS
EST.
ORGANISM
Xenopus laevis (African clawed frog)
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 688)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L.,
Pape, D., Martin, J., Wyllie, T., Underwood, K., Theising, B., Bowers, Y.,
Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Other ESTs: dag55b10.y1
CONTACT: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, E. ellefroid, and A.M. Zorn
(Wellcome/CRC Institute). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
High quality sequence stop: 417.
FEATURES
Location/Qualifiers
1..688
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4785354"
/tissue_type="egg, subtracted by stage 19-26 animal cap"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Wellcome CRC PRN3 St19 26 egg animal cap"
/notes="Vector: pBSRN3; Site 1: NotI; Site 2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subtracted
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute). "
ORIGIN
Alignment Scores:
Pred. NO.: 2.69e-15 Length: 688
Score: 560.50 Matches: 116
Percent Similarity: 61.14% Conservative: 13
Best Local Similarity: 54.98% Mismatches: 25
Query Match: 26.40% Indels: 57
DB: 12 Gaps: 3
US-10-087-080-32 (1-402) x BI443539 (1-688)

```

```

Qy 117 ProLysProPofYrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGly 136
Db 688 CCAAACTTCCTTATTCGTACATTGATAGCCATGGCTACGAGGAGCTCGCCAGT 629
Qy 137 GlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArg 156
Db 628 GCGCCCTGACACTGGCAGAGATCAATGACTATTCATGAAGAATTCCTCTTTCAGG 569
Qy 157 GlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPhe 176
Db 568 GGTAGCTACACTGGCTGGAGGAAGTCTAGTGGACACATCTCTCCCTCAATGACTGTTT 509
Qy 177 ValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsn 196
Db 508 GTCAAGTTCACGGGACCCATCTAGGCCATGGGCAAGATATTAATTACTGATGCTGAAC 449
Qy 197 ProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSer 216
Db 448 CCAAACTGAGTACACCTTTCAGATGGGTCTTCAGACGACAGAGGAGGAGCTAAAC 389
Qy 217 HisArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla 236
Db 388 AGGTCACCAAGTGC-----CTCAAGACGAGGATCTGCAGGACTT----- 347
Qy 237 AlaProProAlaProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGln 256
Db 346 -----GCAGAACAGCAGCAGCAAAATGATGAACCACTAAAGCA 308
Qy 257 GluGluArgAlaSerProAlaGly----- 264
Db 307 AAACAAGAGTGTTCCTCCCTCATCCAGCCGGTGTATCATGGCACATATAAGCACTTCATCC 248
Qy 264 ----- 264
Db 247 TCCTCTACAACTCTCTCTCCACAGGCTACCAAGAAACCAACTCGGGGACCAAGTTC 188
Qy 265 -----LysPheSerSerSerPheAlaIle 272
Db 187 TCCAGCAGGTCTACCAAGAAACCAACTCGGGGACCAAGTTCCTCAGCTCTTCGCAAT 128
Qy 273 AspSerIleLeuArgLysProPheArgSerArg 283
Db 127 GAAGTATCTTGAGTAAACCTTTCAGAGGAGA 95

RESULT 15
CD282719 542 bp mRNA linear EST 23-MAY-2003
LOCUS G38935.47 NCI CGAP_Zemb2 Danio rerio cDNA clone IMAGE:6519563 5',
DEFINITION mRNA sequence.
ACCESSION CD282719
VERSION CD282719.1 GI:31060495
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 542)
Amundsen, C., Cachuela, N., Chen, F., Cheung, L.M., Chong, A.,
Murray, L., Oliva, J., Park, C., Reyes, J., Yungen, J. and Swimmer, C.
Expressed sequence tags from NCI CGAP_Zemb2, a Danio rerio
embryonic library
Unpublished (2003)
Contact: Chen F.
Exelixis, Inc.
170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
Tel: 650 837 7000
Fax: 650 837 8300
Email: fchen@exelixis.com
DNA Sequencing by: Exelixis, Inc. Clone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: 14103 row: N column: 11

```

Job time : 2587 secs

FEATURES High quality sequence stop: 542.  
Location/Qualifiers  
1..542  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:6519563"  
/tissue\_type="embryo"  
/lab\_host="DH10B (TI-resistant)"  
/clone\_lib="NCI CGAP ZEmb2"  
/notes="Vector: PCMV-SpOFT6.1; Site 1: EcoRV; Site 2: NotI;  
cloned unidirectionally. Primer: Oligo dt. Average insert  
size 2 kb. Constructed by J. Wang (Research Genetics,  
Invitrogen Corp) from tissue donated by L. Zon (Harvard  
University). Note: this is a NCI CGAP Library."

source

## ORIGIN

Alignment Scores: 1.29e-13 Length: 542  
Pred. No.: 519.00 Matches: 112  
Score: 519.00  
Percent Similarity: 62.19% Conservative: 13  
Best Local Similarity: 55.72% Mismatches: 24  
Query Match: 24.45% Indels: 52  
DB: 14 Gaps: 5

US-10-087-080-32 (1-402) x CD282719 (1-542)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
DB 96 ATGAAGTTGGAGGTTTC-----TCTGGAGTCGGCTTGTGGACAAGCGCGTGGACCTG 149  
QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40  
DB 150 TGCAGTGACATGGATCGCAACGTGGTCTCACCCTCTCCACG-----GAGGAGGAATTG 203  
QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60  
DB 204 GGCTCGGATGGAGACTGCTCAGCAAC----- 230  
QY 61 GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGluAlaIleProAlaAla 80  
DB 231 -----AGTCAGGACCGCGTGCCCT----- 251  
QY 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaGlyProGlyAlaGly 100  
DB 252 -----GTTCCGGACGG----- 263  
QY 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120  
DB 264 -----AAGCGAACCACATATACGCTAGACCCAAACCACT 299  
QY 121 TyrSerTyrIleAlaIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140  
DB 300 TACTCTTACATTCATTTATGCAATGGCCATACGGATTCCAACACCGCTCGTCTTACA 359  
QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160  
DB 360 CTTGCTGAATAAACCAATACITTAATGAAAAATTCCTCGTTTTCGGGGCAGCTACACG 419  
QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180  
DB 420 GGGTGGCGAAACTCTCTACGGCATACTTGTCAATAACGACTGCTTTTAAAGTGTG 479  
QY 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlu 200  
DB 480 CGAGATCCCTCGGTCGTTGGGCAAGACCATTTATGATGTTGNACCCACACAGCGAG 539  
QY 201 Tyr 201  
DB 540 TAC 542

Search completed: March 31, 2004, 08:38:08

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 30, 2004, 22:00:37 ; Search time 4678 Seconds  
(without alignments)  
11201.738 Million cell updates/sec  
Title: US-10-087-080-31  
Perfect score: 1209  
Sequence: 1 atgaagcttgagggtgttgt.....tggagacgtcttagcttga 1209

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0  
Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb\_ba\*
- 2: gb\_hg\*
- 3: gb\_in\*
- 4: gb\_om\*
- 5: gb\_ov\*
- 6: gb\_pat\*
- 7: gb\_ph\*
- 8: gb\_pl\*
- 9: gb\_pr\*
- 10: gb\_ro\*
- 11: gb\_sts\*
- 12: gb\_sy\*
- 13: gb\_un\*
- 14: gb\_vi\*
- 15: gb\_vl\*
- 16: em\_fun\*
- 17: em\_hum\*
- 18: em\_in\*
- 19: em\_mus\*
- 20: em\_om\*
- 21: em\_or\*
- 22: em\_ov\*
- 23: em\_pat\*
- 24: em\_ph\*
- 25: em\_pl\*
- 26: em\_ro\*
- 27: em\_sts\*
- 28: em\_un\*
- 29: em\_vi\*
- 30: em\_hg\_hum\*
- 31: em\_hg\_inv\*
- 32: em\_hg\_other\*
- 33: em\_hg\_mus\*
- 34: em\_hg\_pln\*
- 35: em\_hg\_rod\*
- 36: em\_hg\_mam\*
- 37: em\_hg\_vrt\*
- 38: em\_sy\*
- 39: em\_hgto\_hum\*
- 40: em\_hgto\_mus\*
- 41: em\_hgto\_other\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1208	99.9	1448	9	AF153341 Homo sapi
2	1183.8	97.9	62761	9	AL499606 Human DNA
3	1180.6	97.7	2322	9	BC053850 Homo sapi
4	1180.6	97.7	4109	9	AF225950 Homo sapi
5	781.6	64.6	1513	10	AF154426 Mus muscu
6	781.6	64.6	2046	10	BC047155 Mus muscu
7	781.6	64.6	187709	10	AL589738 Mouse DNA
8	773.4	64.5	4763	10	AF010405 Mus muscu
9	773.4	64.0	1576	10	AF153193 Rattus ra
10	773.4	64.0	195184	2	AC119497 Rattus ra
11	768.4	63.6	182152	2	AC067929 Rattus no
12	717.8	59.4	1760	10	RATHFH1 Homo sapi
13	653	54.0	178168	2	AC012350 Homo sapi
14	545.4	45.1	74969	2	AC016269 Homo sapi
15	520.8	20.7	182152	2	AC067929 Homo sapi
16	227.8	18.8	2026	5	GGU47276 Gallus gall
17	211.8	17.5	182835	10	AC121839 Mus muscu
18	209.4	17.3	173073	9	AC098522 Homo sapi
19	207	17.1	2437	10	MUSBR21F Mus musculu
20	201.4	16.7	2272	9	HSU59832 Human trans
21	201.4	16.7	5181	6	AX336763 Sequence
22	201.4	16.7	5181	9	HSU59831 Human trans
23	199.6	16.5	2225	10	MMU41047 Human DNA
24	199.6	16.5	2305	10	AF067421 Mus muscu
25	197.8	16.4	2807	10	AF067421 Mus muscu
26	196.8	16.3	62848	5	BSX10653 Zebrafish
27	195.2	16.1	160796	9	AL499604 Human DNA
28	195	16.1	1446	9	AF197560 Homo sapi
29	195	16.1	142728	9	HSU59831 Human trans
30	194.6	16.1	3510	9	HSU59831 Human trans
31	193.4	16.0	22741	2	AC127771 Rattus no
32	193.4	16.0	286524	2	AC105802 Rattus no
33	193	16.0	2011	9	AF275722 Homo sapi
34	193	16.0	106239	9	AL607122 Human DNA
35	192.8	15.9	186734	10	EX005053 Mouse sapien
36	192	15.9	1673	9	HSTIR2 Y13386 Homo sapien
37	189.6	15.7	230905	2	AC132180 Rattus no
38	188.8	15.6	2606	10	MMU457047 Rattus no
39	188.8	15.6	217762	10	AL806523 Mouse DNA
40	188.6	15.6	232988	2	AC095712 Rattus no
41	188	15.6	4259	9	AF042832 Homo sapi
42	187.8	15.5	153604	2	AC022754 Homo sapi
43	187.6	15.5	1860	6	AR083460 Sequence
44	186.6	15.4	185539	2	AC018349 Homo sapi
45	184.4	15.3	1965	6	AR274884 Sequence

ALIGNMENTS

RESULT 1  
AF153341  
LOCUS Homo sapiens winged helix/forkhead transcription factor (HFHL)  
DEFINITION Homo sapiens winged helix/forkhead transcription factor (HFHL)  
gene, complete cds.  
ACCESSION AF153341  
VERSION AF153341.1  
KEYWORDS GI:8489092  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1448)  
AUTHORS Hong H.K., Noveroske J.K., Headon D.J., Liu T., Sy M.S.,  
Justice M.J. and Chakravarti A.

TITLE The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice  
 JOURNAL Genesis 29 (4): 163-171 (2001)  
 MEDLINE 21207067  
 PUBMED 11309849  
 REFERENCE 2 (bases 1 to 1448)  
 AUTHORS Hong, H.-K., Noveroske, J.K., Justice, M.J. and Chakravarti, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-MAY-1999) Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4955, USA  
 FEATURES Location/Qualifiers  
 source 1..1448  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6p25"  
 <233..>1441  
 /gene="HPH1"  
 <233..>1441  
 /gene="HPH1"  
 /product="winged helix/forkhead transcription factor"  
 233..1441  
 /gene="HPH1"  
 /codon\_start=1  
 /product="winged helix/forkhead transcription factor"  
 /protein\_id="AA075586.1"  
 /db\_xref="GI:849093"  
 /translation="MKLEVPRAAHGDKGSLGAGSDAPSPLSAGDSDLSGSDG  
 DCAKPSAGAGDNDGDEOSAGSGGAEALPARAAVAUAEAGAGAGGAGGA  
 GSEGAARKYTRPFPYSYIALIMARDAGGLTLAEINELMGKFFPRGSYT  
 GWRNSVHNSLNDGDFVKVLRDPSRWGKNDYMLNPNSEYTFADGVPFRKRLSHR  
 APVAPGLRPEEAPGLFAPAPPAAPAPSPRMSPAROBERASPAGKSSPAIDSL  
 RKFRRLDTPFTLQWGAAPCPPLPAPFALLPAPCPALLPLCYAGAGEPARLG  
 ARAEVPTAPPILLAPLPAAPAKPLRGPAAAGAHLYCPRLRPLAALQALVRRGPH  
 LSPVETLLA"

gene

mRNA

CDS

ORIGIN

Query Match 99.9%; Score 1208; DB 9; Length 1448;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-120;  
 Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 ATGAAGTTGAGGTGTTCTGCTCCCTCGCGCGGCGCCAGCGGAGCAAGCAGGCGAGTACCTG 60  
 233 ATGAAGTTGAGGTGTTCTGCTCCCTCGCGCGGCGCCAGCGGAGCAAGCAGGCGAGTACCTG 292  
 61 GAGGCGCGGCGGCGGAGCGCGCTGCTCCCTGCTGCGCGGCGGAGCAGCTCCCTG 120  
 293 GAGGCGCGGCGGCGGAGCGCGCTGCTCCCTGCTGCGCGGCGGAGCAGCTCCCTG 352  
 121 GGCTCAGATGGGACTCGCGGCGCAAGCCGCTCCGCGGCGGCGGCGGCGGCGGAGTACGAG 180  
 353 GGCTCAGATGGGACTCGCGGCGCAAGCCGCTCCGCGGCGGCGGCGGCGGCGGAGTACGAG 412  
 181 GCGCAGCGGCGAAGAGATGCGGAGCGGCGGCGGCGGCGGAGAGCGATCCCGGAGCA 240  
 413 GCGCAGCGGCGAAGAGATGCGGAGCGGCGGCGGCGGCGGAGAGCGATCCCGGAGCA 472  
 241 GCTGCTGAGCGGTGCTGCGGAGCGGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 300  
 473 GCTGCTGAGCGGTGCTGCGGAGCGGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 532  
 301 GCGCGGCGGAGCGGCGGAGGTGCGAGCGAGCAAGCGCATATACGCGGCGGCGGCGGCGGCG 360  
 533 GCGCGGCGGAGCGGCGGAGGTGCGAGCGAGCAAGCGCATATACGCGGCGGCGGCGGCGGCG 592  
 361 TACTCGTACATCGGCTCATCGCATCGGCGATCGGCGATCGGCGGCGGCGGCGGCGGCGGCG 420  
 593 TACTCGTACATCGGCTCATCGCATCGGCGATCGGCGATCGGCGGCGGCGGCGGCGGCGGCG 552  
 421 CTGCGGAGATCAACAGATCACTCATGCGGCAAGTTCCTTTTTCGCGGCGGAGTACAG 480  
 553 CTGCGGAGATCAACAGATCACTCATGCGGCAAGTTCCTTTTTCGCGGCGGAGTACAG 712

QY 481 GGTCGGCAACTCCGTGCGGCACAACTTTGCTCAACGACTGCTTCTGTCGAAGTGTCTG 540  
 Db 713 GGTCGGCAACTCCGTGCGGCACAACTTTGCTCAACGACTGCTTCTGTCGAAGTGTCTG 772  
 QY 541 CGGAGCCCTCGCGGCGCTGGGGCAAGGAGCAACTACTGATGCTCAACCCCAACAGCGAG 600  
 Db 773 CGGAGCCCTCGCGGCGCTGGGGCAAGGAGCAACTACTGATGCTCAACCCCAACAGCGAG 832  
 QY 601 TACACCTTCGCGAGGCGGTCTTTCGCGCGCGCGCGCAAGCGCTTACGACACCGCGCGCG 660  
 Db 833 TACACCTTCGCGAGGCGGTCTTTCGCGCGCGCGCGCAAGCGCTTACGACACCGCGCGCG 892  
 QY 661 GTCCCGCGCGCGGCTGCGGCGCGGAGGAGGCGCGCGCGCTCCGCGCGCGCGCGCGCGCG 720  
 Db 893 GTCCCGCGCGCGGCTGCGGCGCGGAGGAGGCGCGCGCGCTCCGCGCGCGCGCGCGCGCG 952  
 QY 721 GCGCGCGCGCGCGGCTGCGCGCGCGATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
 Db 953 GCGCGCGCGCGCGGCTGCGCGCGCGATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1012  
 QY 781 AGCCCGCGCGGCAAGTCTTCCAGCTCTTCCGCAATCGACAGCATCTTCCGCAAGCCCTTC 840  
 Db 1013 AGCCCGCGCGGCAAGTCTTCCAGCTCTTCCGCAATCGACAGCATCTTCCGCAAGCCCTTC 1072  
 QY 841 CGCAGCGCTCGCTCAGGAGACACGCGCGCGCGCGCGCGCTTCACTGCGGCGCGCGCGCGCG 900  
 Db 1073 CGCAGCGCTCGCTCAGGAGACACGCGCGCGCGCGCGCGCTTCACTGCGGCGCGCGCGCGCG 1132  
 QY 901 TGCGCGCGCTCGCGGCTTCCCGCGCTTCCCGCGCGCGCGCGCGCGCTTCACTGCGGCGCGCGCGCG 960  
 Db 1133 TGCGCGCGCTCGCGGCTTCCCGCGCTTCCCGCGCGCGCGCGCGCGCTTCACTGCGGCGCGCGCGCG 1192  
 QY 961 CGCTCTGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGCGCGCGCGCGCGCGCGAGTG 1020  
 Db 1193 CGCTCTGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGCGCGCGCGCGCGCGCGAGTG 1252  
 QY 1021 CCACCGCGCGCGCGCGCTTCTGCTGCTTCACTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1080  
 Db 1253 CCACCGCGCGCGCGCGCTTCTGCTGCTTCACTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1312  
 QY 1081 CTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCTTCACTTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1140  
 Db 1313 CTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCTTCACTTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1372  
 QY 1141 CTCGAGGCGCGCTTACTGCGCGCTTCCCGCGCGCGCGCGCGCTTCACTTCCCGCGCGCGCGCGCGCG 1200  
 Db 1373 CTCGAGGCGCGCTTACTGCGCGCTTCCCGCGCGCGCGCGCGCTTCACTTCCCGCGCGCGCGCGCGCG 1432  
 QY 1201 CTAGCTTGA 1209  
 Db 1433 CTAGCTTGA 1441

RESULT 2  
 AL499606 Human DNA sequence from clone Rp11-13J16 on chromosome 6, complete  
 LOCUS sequence.  
 DEFINITION AL499606  
 ACCESSION AL499606.18 GI:17973944  
 VERSION HTG.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 MAMMALIA; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 62761)  
 Corby, N.  
 Direct Submission  
 Submitted (19-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humuqy@sanger.ac.uk  
 On Dec 20, 2001 this sequence version replaced gi:1792904.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 22:00:37 ; Search time 4678 Seconds  
(without alignments)  
11201.738 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaattgaggtgttgtt.....tgagacgtcttagttga 1209

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*\*

1: gb\_ba:\*\*

2: gb\_btg:\*\*

3: gb\_in:\*\*

4: gb\_ov:\*\*

5: gb\_ov:\*\*

6: gb\_pat:\*\*

7: gb\_ph:\*\*

8: gb\_pl:\*\*

9: gb\_pr:\*\*

10: gb\_ro:\*\*

11: gb\_ste:\*\*

12: gb\_sy:\*\*

13: gb\_un:\*\*

14: gb\_vi:\*\*

15: em\_ba:\*\*

16: em\_fun:\*\*

17: em\_hum:\*\*

18: em\_in:\*\*

19: em\_mu:\*\*

20: em\_om:\*\*

21: em\_or:\*\*

22: em\_ov:\*\*

23: em\_pat:\*\*

24: em\_ph:\*\*

25: em\_pi:\*\*

26: em\_ro:\*\*

27: em\_sts:\*\*

28: em\_un:\*\*

29: em\_vi:\*\*

30: em\_htg\_hum:\*\*

31: em\_htg\_inv:\*\*

32: em\_htg\_other:\*\*

33: em\_htg\_mus:\*\*

34: em\_htg\_pln:\*\*

35: em\_htg\_rod:\*\*

36: em\_htg\_mam:\*\*

37: em\_htg\_vrt:\*\*

38: em\_sy:\*\*

39: em\_htgo\_hum:\*\*

40: em\_htgo\_mus:\*\*

41: em\_htgo\_other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1208	99.9	1448	9	AF153341	AF153341 Homo sapi
2	1183.8	97.9	62761	9	AL499606	AL499606 Human DNA
3	1180.6	97.7	2322	9	BC053850	BC053850 Homo sapi
4	1180.6	97.7	4109	9	AF225950	AF225950 Homo sapi
5	781.6	64.6	1513	10	AF154426	AF154426 Mus muscu
6	781.6	64.6	2046	10	BC047155	BC047155 Mus muscu
7	781.6	64.6	187709	10	AL585738	AL585738 Mouse DNA
8	780	64.5	4763	10	AF010405	AF010405 Mus muscu
9	773.4	64.0	1576	10	AF153193	AF153193 Rattus ra
10	768.4	63.6	182152	2	AC011947	AC011947 Rattus no
11	768.4	63.6	182152	2	AC067929	AC067929 Homo sapi
12	717.8	59.4	1760	10	BATHPH1	BATHPH1 Rattus norv
13	553	54.0	178168	2	AC012350	AC012350 Homo sapi
14	545.4	45.1	74969	2	AC016269	AC016269 Homo sapi
15	250.8	20.7	182152	2	AC067929	AC067929 Homo sapi
16	227.8	18.8	2026	5	GGU47276	U47276 Gallus gall
17	211.8	17.5	183835	10	AC121839	AC121839 Mus muscu
18	209.4	17.3	173073	9	AC099522	AC099522 Homo sapi
19	207	17.1	2437	10	MUSBP2TF	L38607 Mus musculu
20	201.4	16.7	2272	9	HSU59832	U59832 Human trans
21	201.4	16.7	5181	6	AX336763	AX336763 Sequence
22	201.4	16.7	5181	9	HSU59831	U59831 Human trans
23	199.6	16.5	2225	10	MMU41047	U41047 Mus musculu
24	199.6	16.5	2305	10	AF067421	AF067421 Mus muscu
25	197.8	16.4	2807	10	RNTT2PRT	Y11321 R.norvegicu
26	196.8	16.3	69848	5	BX510653	BX510653 Zebrafish
27	195.2	16.1	160796	9	AL499604	AL499604 Human DNA
28	195	16.1	1446	9	AF197560	AF197560 Homo sapi
29	195	16.1	142728	9	HSU792G4	U89995 Human DNA
30	194.6	16.1	3510	9	HSU89995	U89995 Human DNA b
31	193.4	16.0	227341	2	AC127771	AC127771 Rattus no
32	193.4	16.0	286524	2	AC105802	AC105802 Rattus no
33	193	16.0	2011	9	AF275722	AF275722 Homo sapi
34	193	16.0	108239	9	AL607122	AL607122 Human DNA
35	192.8	15.9	186734	10	BX005053	BX005053 Mouse DNA
36	192	15.9	1673	9	HSTIF2	Y13386 Homo sapien
37	189.6	15.7	230905	2	AC132180	AC132180 Rattus no
38	188.8	15.6	2606	10	MMU457047	U457047 Mus muscu
39	188.8	15.6	217762	10	AL806523	AL806523 Mouse DNA
40	188.6	15.6	232988	2	AC095712	AC095712 Rattus no
41	188	15.6	4259	9	AF042832	AF042832 Homo sapi
42	187.8	15.5	153604	2	AC022754	AC022754 Homo sapi
43	187.6	15.5	1860	6	AR083460	AR083460 Sequence
44	186.6	15.4	185539	2	AC018349	AC018349 Homo sapi
45	184.4	15.3	1965	6	AR274884	AR274884 Sequence

ALIGNMENTS

RESULT 1  
AF153341  
LOCUS AF153341 1448 bp DNA linear PRI 24-AUG-2001  
DEFINITION Homo sapiens winged helix/forkhead transcription factor (HPH1)  
ACCESSION AF153341  
VERSION AF153341.1 GI:8489092  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1448)  
HONG,H.K., Noveroske,J.K., Headon,D.J., Liu,T., SY,M.S.,  
Justice,M.J. and Chakravarti,A.

TITLE The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satini mice  
 JOURNAL Genesis 29 (4), 163-171 (2001)  
 MEDLINE 21207067  
 PUBMED 11309849  
 REFERENCE 2 (bases 1 to 1448)  
 AUTHORS Hong, H.-K., Noveroske, J.K., Justice, M.J. and Chakravarti, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-MAY-1999) Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4955, USA

## FEATURES

Location/Qualifiers  
 1..1448  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6p25"  
 <233..>1441  
 /gene="HFH1"  
 <233..>1441  
 /gene="HFH1"  
 /product="winged helix/forkhead transcription factor"  
 233..1441  
 /gene="HFH1"  
 /codon\_start=1  
 /product="winged helix/forkhead transcription factor"  
 /protein\_id="AAF75586.1"  
 /db\_xref="GI:8489093"  
 /translation="MKLEVFVRAAHGDKQSDLEAGGSDAPSLSAAGDLSGDS  
 DCAKPSAGGARDTQGGEGSAGGPGAEPAIAAAVAVAGRAAGPAGGA  
 GSGEARKPYTRRKPYSYIALAMAIRSAGRLTAEINYLWGKFFFRGSYT  
 GWNSVRLNLSLNDGCFVKLRDPSPKGNKYNMLNPNSEYTFADGVPFRRLKLSH  
 APVPAPGRPEELPGLPAAPPAPAPASPRMRSPARQERASPAKFPSSFLDSIL  
 RPFPSRLRLDAPGTLTQWGAAPCPPLPAFPALLPAAPCPALLPLCAYGAGEPARLG  
 AREAEVPTAPTPLLAPLPAAPAKPLRGAAGAHLYCPLRLPAALQAALVRRPGH  
 LSPVETILLA"

## ORIGIN

Query Match 99.9%; Score 1208; DB 9; Length 1448;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-120;  
 Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTTGGAGTGTTCCTCGCGCGGCCACGGGACAAAGCAGGCGAGTACCTG 60  
 DB 233 ATGAAGTTGGAGTGTTCCTCGCGCGGCCACGGGACAAAGCAGGCGAGTACCTG 292

QY 61 GAGGCGCGGCGGAGAGACGCGCGTCCCGCTGTTCGGCGGCGGAGACGACTCCCTG 120  
 DB 293 GAGGCGCGGCGGAGAGACGCGCGTCCCGCTGTTCGGCGGCGGAGACGACTCCCTG 352

QY 121 GGCTCAGATGGGACTCGCGGCCCAAGCGTCCGCGGCGGCGGCGGCGGAGATACGAG 180  
 DB 353 GGCTCAGATGGGACTCGCGGCCCAAGCGTCCGCGGCGGCGGCGGCGGAGATACGAG 412

QY 181 GGCAGCGGCGAACAAGATGTCGGAGGCGGCGGCGGCGGCGGAGGCGGATCCCGGAGCA 240  
 DB 413 GGCAGCGGCGAACAAGATGTCGGAGGCGGCGGCGGCGGCGGAGGCGGATCCCGGAGCA 472

QY 241 GTGTGTGACGCGTGTGTCGAGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCG 300  
 DB 473 GTGTGTGACGCGTGTGTCGAGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCG 532

QY 301 GCGCGGCGGAGCGGCGAGGCGTGCAGCGAGAGCCATATAGCGGCGGCGGCGGCGGCGGCG 360  
 DB 533 GCGCGGCGGAGCGGCGAGGCGTGCAGCGAGAGCCATATAGCGGCGGCGGCGGCGGCGGCG 592

QY 361 TACTGTATATCGCGTCTATCCCATGCGCATCCCGCATCCCGCGGCGGCGGCGGCGTACG 420  
 DB 593 TACTGTATATCGCGTCTATCCCATGCGCATCCCGCATCCCGCGGCGGCGGCGGCGTACG 652

QY 421 CTGGCGGAGATCAACGAGTACCTCATGCGAGTTCCTCTTTTCCGGGCGGAGTACACG 480  
 DB 653 CTGGCGGAGATCAACGAGTACCTCATGCGAGTTCCTCTTTTCCGGGCGGAGTACACG 712

QY 481 GGCTGGCGCAACTCCGTCGCGCAAACTTTTCTGCTCAAGACTGCTTCTGTCGAAGTGTCTG 540  
 DB 713 GGCTGGCGCAACTCCGTCGCGCAAACTTTTCTGCTCAAGACTGCTTCTGTCGAAGTGTCTG 772

QY 541 CGCGACCCCTCGCGGCGCTGGGCGCAAGGACAACTACTTGGATGCTCAACCCCAACAGCGAG 600  
 DB 773 CGCGACCCCTCGCGGCGCTGGGCGCAAGGACAACTACTTGGATGCTCAACCCCAACAGCGAG 832

QY 601 TACACTTTCGCCGACGGGGTCTTCGCGCGCGCGCGCAAGCGCTTCAGCCACCGCGGCGCG 660  
 DB 833 TACACTTTCGCCGACGGGGTCTTCGCGCGCGCGCGCAAGCGCTTCAGCCACCGCGGCGCG 892

QY 661 GTCCCGCGCGCGCGGCTGGCGCGCGGAGGCGCGCGGCGCTCCCGCGCGCGCGCGCGCG 720  
 DB 893 GTCCCGCGCGCGCGGCTGGCGCGCGGAGGCGCGCGGCGCTCCCGCGCGCGCGCGCGCG 952

QY 721 GCGCGCGCGCGCGCGCTTCGCGCGCGCATGCGCTTCGCTCGCGCGCGCGCGCGCGCGCGCG 780  
 DB 953 GCGCGCGCGCGCGCGCTTCGCGCGCGCATGCGCTTCGCTCGCGCGCGCGCGCGCGCGCGCG 1012

QY 781 AGCCCGCGCGCGCAAGTCTCCAGCTCTCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 840  
 DB 1013 AGCCCGCGCGCGCAAGTCTCCAGCTCTCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1072

QY 841 CGCAGCGCTCGCTTCAGGGACACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
 DB 1073 CGCAGCGCTCGCTTCAGGGACACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1132

QY 901 TGCCTGGCGCTGCGCGCGCTTCGCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 960  
 DB 1133 TGCCTGGCGCTGCGCGCGCTTCGCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 1192

QY 961 CCGCTCTGCGCGTACG 1020  
 DB 1193 CCGCTCTGCGGTACG 1252

QY 1021 CCACCGACCGCGCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 1080  
 DB 1253 CCACCGACCGCGCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 1312

QY 1081 CTCGAGGCG 1140  
 DB 1313 CTCGAGGCG 1372

QY 1141 CTCGAGGCGCGCTTACTTCG 1200  
 DB 1373 CTCGAGGCGCGCTTACTTCG 1432

QY 1201 CTAGCTTGA 1209  
 DB 1433 CTAGCTTGA 1441

## RESULT 2

AL499606 62761 bp DNA linear PRI 19-DEC-2001  
 LOCUS Human DNA sequence from clone Rp11-13116 on chromosome 6, complete  
 DEFINITION sequence.  
 ACCESSION AL499606  
 VERSION AL499606.18 GI:17973944  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (Bases 1 to 62761)  
 Corby, N.  
 Direct Submission  
 Submitted (19-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Dec 20, 2001 this sequence version replaced gi:17902904.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>

RP11-13J16 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone

RP11-13J16 It may be shorter because we sequence overlapping

sections only once, except for a short overlap.

The true left end of clone RP11-13J16 is at 1 in this sequence. The

true left end of clone RP4-668J24 is at 60762 in this sequence. The

true right end of clone RP1-116B8 is at 4171 in this sequence.

Location/Qualifiers

1..62761

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="6"

/clone="RP11-13J16"

/clone\_lib="RP11-11.1"

17171..17234

/note="Single clone region. Sequence from reads from a

short insert library derived from a single pUC clone.

Restriction digest data confirm the assembly."

17634..17641

/note="Single clone region. Sequence from reads from a

short insert library derived from a single pUC clone.

Restriction digest data confirm the assembly."

21165..21540

/note="Single clone region. Sequence from reads from a

short insert library derived from a single pUC clone.

Assembly consistent with restriction digest data."

21466

/note="Tandem repeat. Forced join. Gap size estimated to

be approximately 430bp by restriction digest data."

23824..24045

/note="Single clone region. Sequence from reads from a

short insert library derived from a single pUC clone.

Restriction digest data confirm the assembly."

60092

/note="Tandem repeat. Forced join. Gap size estimated to

be approximately 120bp by restriction digest data."

# ORIGIN

Query Match 97.9%; Score 1183.8; DB 9; Length 62761;  
 Best Local Similarity 99.1%; Pred. No. 1.2e-118;  
 Matches 1201; Conservative 0; Mismatches 8; Indels 3; Gaps 1;  
 1 ATGAAGTTGGAGGTTTCCTCCCTCGCGGGCCCAAGCGGACAGCGGCGAGTGCAGCTG 60  
 16945 ATGAAGTTGGAGGTTTCCTCCCTCGCGGGCCCAAGCGGACAGCGGCGAGTGCAGCTG 17004  
 61 GAGGGCGGGCGGCGAGCGAGCGGCGCTCCCGCTGTCTCGGGCGGGAGACGATCCCTG 120

Db	17005	GAGGGCGGGCGGCGAGCGAGCGGCGCTCCCGCTGTCTCGGGCGGAGACGACTCCCTG	17064
Qy	121	GGTCAAGATGGGACATGCGCGGCGCAA---GCCGTTCGGCGGGCGGGCGGCGCAGATACG	177
Db	17065	GGTCAAGATGGGACATGCGCGGCGCAAACAGCCCGCGCGGGCGGCGGCGCAGATACG	17124
Qy	178	CAGGGCGAGCGGCGAAACAGAGTGGGGAGGCGGGCGGGCGGCGGAGGAGGATCCCGGCA	237
Db	17125	CAGGGCGAGCGGCGAAACAGAGTGGGGAGGCGGGCGGGCGGCGGAGGAGGATCCCGGCA	17184
Qy	238	GCAGCTGCTGCACGCGGTGGTGGCGAGGCGGCGGCGGGCGGGCGGCGGCGGCGG	287
Db	17185	GCAGCTGCTGCACGCGGTGGTGGCGAGGCGGCGGCGGGCGGGCGGCGGCGGCGG	17244
Qy	298	GGGGCGGGGAGCGGCGAGGGTGCACGACGACGACATATACGCGGGCGGCGGCGGCGG	357
Db	17245	GGGGCGGGGAGCGGCGAGGGTGCACGACGACGACATATACGCGGGCGGCGGCGGCGG	17304
Qy	358	CCCTACTCGTACATCGCGCTCATCGCATGGCCATCCGGGACATCGGCGGGCGGCGGCTTG	417
Db	17305	CCCTACTCGTACATCGCGCTCATCGCATGGCCATCCGGGACATCGGCGGGCGGCGGCTTG	17364
Qy	418	ACCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTTCGGCGGCGGCTAC	477
Db	17365	ACCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTTCGGCGGCGGCTAC	17424
Qy	478	ACGGGCTGGCGCAACTCCGTGGCGCACAACTTTTCCTCAACGACATGTTCTGTCAAGGTG	537
Db	17425	ACGGGCTGGCGCAACTCCGTGGCGCACAACTTTTCCTCAACGACATGTTCTGTCAAGGTG	17484
Qy	538	CTCGGCGACCTTCGGCGGCTTGGGGCGAAGCACTACTGATGCTCAACCCCAACAGC	597
Db	17485	CTCGGCGACCTTCGGCGGCTTGGGGCGAAGCACTACTGATGCTCAACCCCAACAGC	17544
Qy	598	GAGTACACTTTCGCGACGCGGCTTTCGCGCGCGCGCAAGCGCTCAGGCGGCGGCGG	657
Db	17545	GAGTACACTTTCGCGACGCGGCTTTCGCGCGCGCGCAAGCGCTCAGGCGGCGGCGG	17604
Qy	658	CCGGTCCCGCGCGCGGCTGGGGCGGAGGAGGCGGCGGCGCTCCCGCGGCGGCGGCGG	717
Db	17605	CCGGTCCCGCGCGCGGCTGGGGCGGAGGAGGCGGCGGCGCTCCCGCGGCGGCGGCGG	17664
Qy	718	CCCGCGCGCGCGCGCGGCTCGCCCGCGCATGCGCTCCCGCGCGCGCGCGCGGAGCGG	777
Db	17665	CCCGCGCGCGCGCGCGGCTCGCCCGCGCATGCGCTCCCGCGCGCGCGCGCGGAGCGG	17724
Qy	778	GCAGAGCCCGCGGCGCAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTCGGCAAGCCC	837
Db	17725	GCAGAGCCCGCGGCGCAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTCGGCAAGCCC	17784
Qy	838	TTCCGAGCGCTGCGCTCAGGAGACAGGCGCGCGGAGACAGCGTTTCAGTGGGCGCGCG	897
Db	17785	TTCCGAGCGCTGCGCTCAGGAGACAGGCGCGCGGAGACAGCGTTTCAGTGGGCGCGCG	17844
Qy	898	CCCTGCGCGCGCTGCGCGCTTCCCGCGGCTCCTCCCGCGCGCGCGCTCGAGGGGCGCTG	957
Db	17845	CCCTGCGCGCGCTGCGCGCTTCCCGCGGCTCCTCCCGCGCGCGCGCTCGAGGGGCGCTG	17904
Qy	958	CTGCGGCTCTGCGGTACGCGCGGCGGAGCGCGCGCGCTGGGCGGCGGCGGCGGCGG	1017
Db	17905	CTGCGGCTCTGCGGTACGCGCGGCGGAGCGCGCGCGCTGGGCGGCGGCGGCGGCGG	17964
Qy	1018	GTGCCACCGAGCGCGCGCTTCTTCTGTTGACACTTCTCCCGCGCGGCGGCGGCGGCGG	1077
Db	17965	GTGCCACCGAGCGCGCGCTTCTTCTGTTGACACTTCTCCCGCGCGGCGGCGGCGGCGG	18024
Qy	1078	CACTCCGAGGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1137
Db	18025	CACTCCGAGGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	18084
Qy	1138	GCCTCTCAGCGCGCTTAGTCCGNCCTCTCTGGCGCGCACTGTCTGTAACCGGTGAGAGC	1197









[illegible][illegible]

/chromosome="13"  
/clone="RP23-322J11"  
/clone\_lib="RPCI-23"

ORIGIN

Query Match 64.6%; Score 781.6; DB 10; Length 187709;  
Best Local Similarity 81.2%; Pred. No. 9.4e-76;  
Matches 991; Conservative 0; Mismatches 200; Indels 30; Gaps 6;  
QY 1 ATGAATTGGAGTGTTCCTCCCTCGCGGGCCACAGCGGACAGCGAGTACCTG 60  
Db 135706 ATGAATTGGAGTGTTCCTCCACAGCGGACAGCGGACAGCGGAGTATG 135765  
QY 61 GAGGCGGGGCGGAGGAGCGCGGTCCTCCCGTGTTCGGCGGGGAGAGACTCCCTG 120  
Db 135766 GAGGGGGCGGAGGAGCGACGTGACATCTCCACTGTCCGCGGTGGTACACTCTTA 135825  
QY 121 GCTCAGATGGGACTCGCGGCCAA---GCCGTCCGGGGGCGGGCGCCAGAGATACG 177  
Db 135826 GCTCAGAGGGGACTGTGACGCAACAGCCCGCGGGGAGCGGCGCGGGATCTG 135885  
QY 178 CA---GGGCGAGCGGAAACAGAGTGGGAGCGCGGGCGGGCGGAGGCGATCCCG 234  
Db 135886 GAAGGTGGCGGGCGGAGAGGAATTGAGTGGCGGGCGAGCGCCCAAGACGTCGGAG 135945  
QY 235 GCAGCAGCTGCTGACGCGTGTGCGGAGGCGGAGGCGGGCGGGCGGCGGCGG 294  
Db 135946 GCAACTGATG-----ACAGCAAGACGAGGCTCCCGGAGGGCGGTGC 135990  
QY 295 GCGGGCGGCGGGAGCGCGGAGGTGCACGAGCAAGCCATATACGCGGCGGCCAAG 354  
Db 135991 GCGGGCGGCGTGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 136050  
QY 355 CCCCCCTACTGTATATCGCGGTATCGCCATGCGCATCGGATCGCGGGCGGGCGG 414  
Db 136051 CCCCCATCTCTACATCTCTATCGCCATGCGCATCGGATCGCGGGCGGAGCG 136110  
QY 415 TTGACCTGGGAGATCAACGAGTACCTCATGGGCAAGTCCCTTTTCGCGCGCAGC 474  
Db 136111 CTGACACTGGCGAGATCAACGAGTACCTCATGGGCAAGTCCCTTTTCGCGGCGCAGC 136170  
QY 475 TACCGGCTGGCGAACTTCGTTGCGCCACACCTTTTCGCTCAACGATCTTCGTAAG 534  
Db 136171 TACCGGCTGGCGAACTTCGTTGCGCCACACCTTCGCTCAACGATCTTCGTAAG 136230  
QY 535 GTGCTGGGAGCCCTCGCGGCGCTGGGCAAGGACAACTACTGGATGCTCAACCCCAAC 594  
Db 136231 GTGCTGGGAGCCCTCGCGGCGCTGGGCAAGGACAACTACTGGATGCTCAACCCCAAC 136290  
QY 595 AGCGAGTACACCTTCGCGGAGCGGGTCTTCGCGCGCGCGCAAGCGCTCAGCACCGC 654  
Db 136291 AGCGAATACACCTTCGCGGAGCGGGTCTTCGCGCGCGCGCAAGCGCTCAGCACCGC 136350  
QY 655 GCGCGGTCCTCCGCGCGCGGCTCGCGCGGAGGCGCGGCGCTCCCGCGCG---CC 711  
Db 136351 ACCACAGTCTCCGCTCGCGGCTCGCGCGGAGGAGAGCCCAACCGGACTCGCGGACC 136410  
QY 712 CCGCCGCGCGCGCGCGCGCGCTCGCGCGGATCGCTTCGCGCGCGCGCGCGCGGAG 771  
Db 136411 CCGCAGCGCGCGCGCGCGCGCTCGCGCGGATCGCGCGGCTCGCGCGCTCGCGGAG 136470  
QY 772 GAGCGCGCAGCGCGCGGCGGAGTCTTCAGTCTCTTGGCATTCAGAGCATCTCGCG 831  
Db 136471 GAGCGCTCCAGCGCTCGAGAGTCTTCAGTCTCTTGGCATTCAGAGCATCTTCAGC 136530  
QY 832 AAGCCCTTCGCGAGCGCTCGCCTCAGGACACGCGCGCGCGGAGCGGCTTCAGTGGGCG 891  
Db 136531 AAGCCTTTCGAGCGCGCGGAGCGGAGTCTCGGCTCTGGGGTGCAGTACCTCGGGC 136590  
QY 892 GCGCGCGCTCGCGCGCGCTTCGCGCGGCTTCGCGCGGCTTCGCGCGGCGCGCTTCAG 951  
Db 136591 GCGCGCTCGCGCGCGCTTCGCGCGGCTTCGCGCGGCTTCGCGCGGCGCGCGGCTGC 136650

QY 952 GCCTGCTGCGCTCTGCGGTACGCGCGGCGGAGCGCGGCTGGGCGCGCGGAG 1011  
Db 136651 GCTCTGCTACCGCTCTGTCTTACGCGGCAAGGAGCCCTACGCTGTGGGTGCGCGGG 136710  
QY 1012 GCGGAGTGCACCGGACCGCGCGCGCTCTGCTTGCACCTTCCCGGCGGCGGCGGCCCC 1071  
Db 136711 ACCGAGTG---CAGCCCGCGGCGCGCTTGTGTGCGCGCGCTCTCCACCGGGTCCA 136767  
QY 1072 GCAAGCACTCCGAGGCGCGG---CGCGCGGCGGCGGACCTGTACTGCCCCCTCGCG 1128  
Db 136768 GCAAGCACTCCGAGTCCGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCTACGG 136827  
QY 1129 CTGCGCGCACCTCTCAGGCGGCTTAGTCCGNCGTCCTGCGCGCGCACCTGTCTACCCG 1188  
Db 136828 CTGCGCGCACCTCTCAGGCGGCGGCGGCTCGGTCGCGGCTCGGACCTGTCTACCCG 136887  
QY 1189 GTGAGAGCGTCTAGCTTGA 1209  
Db 136888 GTGAGACTCTGCTAGCTTGA 136908  
RESULT 8  
AF010405 4763 bp DNA linear ROD 20-JAN-2000  
LOCUS Mus musculus fork head transcription factor (Hfh-1L) gene, complete cds.  
DEFINITION AF010405 GI:6716869  
ACCESSION AF010405  
VERSION AF010405.2  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 4763)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE Mouse HNF-3/fork head homolog-1-like gene: structure, chromosomal location, and expression in adult and embryonic kidney  
JOURNAL DNA Cell Biol. 17 (8), 679-688 (1998)  
MEDLINE 98392851  
PUBMED 9726250  
REFERENCE 2 (bases 1 to 4763)  
AUTHORS Frank, S. and Zoll, B.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-1997) University of Goettingen, Institute of Human Genetics, Gosslerstr 12d, Goettingen 37073, Germany  
REFERENCE 3 (bases 1 to 4763)  
AUTHORS Pasche, B., Bieller, A. and Zoll, B.  
TITLE Direct Submission  
JOURNAL Submitted (19-JAN-2000) University of Goettingen, Institute of Human Genetics, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany  
REMARK Sequence update by submitter  
COMMENT On Jan 20, 2000 this sequence version replaced gi:2997586.  
FEATURES  
source 1..4763  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="129"  
/db\_xref="taxon:10090"  
/chromosome="13"  
/tissue\_type="kidney"  
/gene="Hfh-1L"  
/note="HNF-3/fork head homologue-1 like"  
/cds="2467..3669"  
/product="fork head transcription factor"  
/gene="Hfh-1L"  
/note="member of HNF-3/fork head transcription factor family"  
/codon\_start=1  
/product="fork head transcription factor"  
/protein\_id="AAC12973.2"

```
/db_xref="GI:6716870"
/translation="MKLEVFVPRAAAHGDKMGLDLEGAGSDVSPSLAAGDSDGLSDG
DCAANSPPAAGSAGLEGGGGRNSGGPSAODPEATDDSRTOASAAAGPAGGVGG
EGARSKPYRRPKPPYSYITALIARDSAGRLTAEINVEYMGKFPFFRGSYTGWR
NSVRNLSLNDFCVLEDPSPSPKWDNYWLNPNSEYTFADGVFRERKLSHRTTV
SASGLRPERAPQAGTPOAAPARSSPIARSPARQBERSSPASKFSSPAIDSLK
PFRSRDGSALGVQIPWGAACPPLIRAYFALLPAAGCALPLPCYAGSEPTLLASR
GTEVQPAAPALLLAPLSTAAAPKFRFPETAGAAHLYCFLRLPTALQAAAAAGPGLHS
YFVETLLA"
misc_feature
2782..3123
/genes="Hfh-1L"
/note="encodes DNA binding domain"

ORIGIN
Query Match 64.5%; Score 780; DB 10; Length 4763;
Best Local Similarity 81.1%; Pred. No. 4.9e-75;
Matches 990; Conservative 0; Mismatches 201; Indels 30; Gaps 6;

QY 1 ATGAAGTTGAGGTGTTGTCCTCCGCGCGGCCACGGGACAAAGCAGGCGAGTGACCTG 60
Db 2467 ATGAANTTGAAGTGTTCCTCCACGCGCAGCCACGGGACAAATGGCAGCGATCTG 2526

QY 61 GAGGCGCGCGGCGGCGACGACGCGCGCTCCCGCTGTCGGCGGCGGAGACGACTCCCTG 120
Db 2527 GAGGCGCGGCGGCGACGACGCGCTCCACTGTCCGCGCGCTGGTGACGACTCCTTA 2586

QY 121 GCCTCAGATGGGACTCGCGCGGCA--GCCGTCCGCGCGGCGGCGGCCACAGATACG 177
Db 2587 GGCTCAGACGGGACTGTGCAGCCACAGCCCGCGCGGCGGAGCGCCCGGATCTG 2646

QY 178 CA---GGGCGACGCGGAACAGAGTGGCGGAGCGCGCGCGCGGCGGAGGCGCATCCG 234
Db 2647 GAAGGTGGCGGCGCGCAGAGGAATTGAGTGGCGGCGCGGCGGCGGCCCAAGCGTCCGGAG 2706

QY 235 GCAGCAGCTCTCAGCGGTGTGGCGGAGGCGCGGAGCGCGGCGCGGCGGCGGCGGCG 294
Db 2707 GCAACTGATG-----ACAGCAGAACGACGCGCTCCCGCGCAGCGGCGCGTGC 2751

QY 295 GCGGCGCGCGCGGCGGCGCGGAGTGGCGGAGCGCGCGCGCGCGCATATACGCGCGCGCCCAAG 354
Db 2752 GCGGCGCGGTGGCGGCGCGGAGGCGCGCGCGCAGCAAGCGGTACACGCGCGCGGCCCAAG 2811

QY 355 CCGCCCTACTGTACATCGCGCTCATCGCATGGCCATCGGACTCGCGGCGGCGGCGGCG 414
Db 2812 CCGCCCTACTGTACATCGCTCTCATCGCATGGCCATCGGACTCGCGGCGGCGGCGGCG 2871

QY 415 TTGACGTGGCGGAGATCAACGAGTACCTCATGGCAAGTTCCTCCCTTTTCCGCGGCGAGC 474
Db 2872 CTGACACTGGCGGAGATCAACGAGTACCTCATGGCAAGTTCCTCCCTTTTCCGCGGCGAGC 2931

QY 475 TACAGGGGTGGCGCACTCCGTGGCGCACAACTTTGGTCAAGACTGCTTCGTCAAG 534
Db 2932 TACAGGGGTGGCGCACTCCGTGGCGCACAACTTCGTCAAGACTGTTTCGTCAAG 2991

QY 535 GTGTGGCGGACCCCTCGCGGCGCTGGGCGAAGCAACTACTGGATGCTCAACCCCAAC 594
Db 2992 GTGTGGCGGACCCCTCGCGGCGCTGGGCGAAGCAACTACTGGATGCTCAACCCCAAC 3051

QY 595 AGCGAGTACACTTGGCCGACGGGTTTCGCGCGCGCGCGCGCGCGCTCCCGCGCG--CC 711
Db 3052 AGCGAATACACTTGGCCGACGGGTTTCGCGCGCGCGCGCGCGCGCTCCCGCGCGCGCG 3111

QY 655 GCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCCGCGCG--CC 711
Db 3112 ACCMAGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3171

QY 712 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 771
Db 3172 CCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3231

QY 772 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 831
Db 3232 GAGCGCTCCAGCGCGTGGCGCAAGTTCTCCAGTCTCTTCCCGATCCGACGATCTTCAGC 3291
```

```
QY 832 AAGCCCTTCGCGAGCGGTCCCTCAGGACACCGGCCCCCGGAGCGAGCTTTCAGTGGGGC 891
Db 3292 AAGCCCTTTTCGCGAGCGGCGCGGAGCTCGGCTCTGGGGGTGCGAGCTTACCTGGGGC 3351

QY 892 GCGCGCGCTTCGCGCGCGCTCCCGGCTTCCCGCGCTTCCCGCGCGCGCGCGCGCTG 951
Db 3352 GCGCTCCCTTCGCGCGCGCTTCGCGGCTATCCCGGCTTCTTCCCGCGCGCGCGGTCG 3411

QY 952 GCGCTGTGTCGCTTCGCGGTACGCGCGGCGGCGAGCGCGCGCGGCTGGCGGCGCGGAG 1011
Db 3412 GCTCTGTCTACGCTCTGTGCTTACGCGGCGAAGCGAGCTTACGCTGTGGGCTCGCGCGG 3471

QY 1012 GCGGAGGTGCCACGCGCGCGCGCTCTCTGCTGACCTCTCCCGCGCGCGCGCGCGCG 1071
Db 3472 ACCGAGTG---CAGCCCGCGGCGCGCTTCTGCTGGCGCGCTCTCCACCGCGGCTCCA 3528

QY 1072 GCCAAGCCACTCGAGGCGCGG---CGCGCGGCGCGCGCGACCTGTACTGCGCGCTCGG 1128
Db 3529 GCCAAGCCATTTCGAGTTCGCGGAGACCGCGCGGCGCGCGACCTGTACTGCGCGCTACG 3568

QY 1129 CTGCGCGCGCGCTTCGAGGCGCGCTTAGTCCGCGCTCTGCGCGCGCGCGCGCGCT 1188
Db 3589 CTGCGCGCGCGCTTCGAGGCGCGCGCGCTGGGTCGCGCTCGGCGCGCGCGCTTCTTAC 3648

QY 1189 GTGGAGACGCTCTCTAGCTTGA 1209
Db 3649 GTGGAGACTCTGCTAGCTTGA 3669

RESULT 9
AF153193 1576 bp DNA linear ROD 24-AUG-2001
LOCUS Rattus rattus winged helix/forkhead transcription factor HFH1
DEFINITION AF153193
ACCESSION AF153193
VERSION AF153193.1 GI:8132282
KEYWORDS Rattus rattus (black rat)
ORGANISM Rattus rattus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1576)
AUTHORS Hong,H.-K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S.,
Justice,M.J. and Chakravarti,A.
TITLE The winged helix/forkhead transcription factor Foxq1 regulates
differentiation of hair in satin mice
JOURNAL Genesis 29 (4), 163-171 (2001)
MEDLINE 21207067
PMID 11309849
REFERENCE 2 (bases 1 to 1576)
AUTHORS Hong,H.-K., Noveroske,J.K., Justice,M.J. and Chakravarti,A.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Genetics, Case Western Reserve University,
10900 Euclid Ave., Cleveland, OH 44106-4955
FEATURES
Source
1. 1576
/organism="Rattus rattus"
/mol_type="genomic DNA"
/db_xref="taxon:10117"
<95..>1297
/gene="Hfh1"
<95..>1297
/gene="Hfh1"
/product="winged helix/forkhead transcription factor HFH1"
95..1297
/gene="Hfh1"
/codon_start=1
/product="winged helix/forkhead transcription factor HFH1"
/protein_id="AAF73234.1"
/db_xref="GI:8132283"
/translation="MKLEVFAPRAAHGDKMGLDLEGAGSDVSPSLAAGDSDGLSDG
```

Qy	1	ATGAAGTTGGAGGTGTTTCGTCCCTCGCGCGCCACCGGGACAAAGCAAGGGCAGTGACCTG	60
Db	95	ATGAAGTTGGAGGTATTTTCGCCACCGCGACGCCACCGGGACAAAGATGGGCAAGTGACCTG	154
Qy	61	GAGGGCGGGGCGCAGACGACGCGCGCTCCCGCTGTCCGCGCGGAGACGACTCCCTG	120
Db	155	GAGGGGCGGGCAGCAGACGAGTGTCGCATCTCCGCTGTCCGGGCTGGCGACGACTCCTTA	214
Qy	121	GGCTCAGATGGGGAATGCGCGGCCAA--GCCGTCCGCGGCGCGCGCGCAGAGATAGC	177
Db	215	GGCTCTGACGGGGAATGTGCAACCCAAAGCCCGCGCGCGGCAAGAGCGCGCTGGATCTG	274
Qy	178	CAGGGCAGCGGACACAGTGTCCGGAGCGGGCGCGGAGGAGGCGATCCCGGCA	237
Db	275	GAAGGGCGGGCGGCAGAGAAATTCAGTGG-----CGGGCGAGCACCCAA	322
Qy	238	GCAGCTGTGTCAAGCGGTGGTGGCGAGGGCGCGAGCGCGGGCGCGGGGCCAGCGCG	297
Db	323	GACGATCCCGAGGTGACCGATGCGCAGCAGAAACGACAGCCTCCCGGTTGGGCGCGTGC	382
Qy	298	GGCGGCGGGGAGCGGGCAGGGTGCACGACAGAGCCATATACGGCGGGCGGCCAGGCC	357
Db	383	GGCAGCGTGGGCGGGGTGAGGGCGCGCAGCAAGCCGTACACGGCGGGCGCCCAAGGCC	442
Qy	358	CCCTACTCTATCATCGCGCTCATGCCATGSCCATCCGCGACTCGGCGGGCGGGCGCTTG	417
Db	443	CCCTACTCTCTCATCGCACTCATCGCATGSCCATCCGCGACTCCGCGGGCGGAGCGCTG	502
Qy	418	ACGCTGGCGGAGATCAACAGGTAACCTCATGGGCAAGTTCCCTTTTTCGCGCGCAGCTAC	477
Db	503	ACGCTGGCGGAGATCAACAGGTACCTCATGGGCAAGTTCCCTTTTTCGCGGGCAGCTAC	562
Qy	478	ACGGGTGCGCAACTCCGTGCGCCACAACTTTTCGTCAACGACTGCTTCGTCAAGGTG	537
Db	563	ACGGGTGCGCAACTCCGTGGGCCAACAACCTCTCGCTCAACGACTGTTTCGTCAAGGTG	622
Qy	538	CTGCGCACTCCCTCGCGCCCTGGGGCAAGCAAACTACTGGATGCTCAACCCCAACAGC	597
Db	623	CTGCGCACTCCCTCGCGCCCTGGGGCAAGCAAAATTAATGGATGCTCAACCCCAACAGC	682
Qy	598	GAGTACACCTTCGCGCAGCGGGTCTTCGCGCGCGCGCAGCGGCTCAGCCACCGCGG	657
Db	683	GAATACACCTTCGCGCAGCGGGTCTTCGCGCGCGCGCAGCGGCTCAGCCACCGGAC	742
Qy	658	CGGTCGCCCGCGCGGGCTGCGGCGCCGAGAGAGCCCGGGCTCCCGCGCG---CGCCG	714
Db	743	ACAGTCTTCGCATCGGGCTACGGCGGAGGAAGCCCAACCGGACCTGGGGGACCCCG	802
Qy	715	CGGCGCGCGCGCGCGCCCGGCTTCGGCCCGCATGTCGCTCGCCCGCGCGCAGAGGAG	774
Db	803	CAGCCCGCGCCACCGCGGGTCTCTCCCCCAATCGCGCGCTCGCCCGCTCGCCAGGAGG	862
Qy	775	CGGCGCAGCCCGCGGGCAAGTTCTCCAGCTTCCTTCGCCATCGACAGCATCTTCGCCAAG	834
Db	863	GGCTCAGCCCGCGGAGCAAGTTCTCAGCTCTCTTCGCCATCGNAGCATCTCTCAGCAAG	922
Qy	835	CCCTTCGCGAGCGCTCGCCTCAGGAGCAACGCGCCCGCGGZACGACTTCAGTGGGCGCGC	894
Db	923	CCGTTTCGAGCGCGCGGACGCGCAGCCCGGCTCTGGGGGTGAGCTTACCTGGAGCGCT	982

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., S., Shen, H., Sanders, W., Savary, G., Scherer, S., Scott, J., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, D., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Yoon, L., Yoon, V., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished  
2 (bases 1 to 195184)  
Worley, K.C.

Direct Submission

Submitted (27-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 195184)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23195197.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GWQI

Center clone name: CH230-521F8

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 158913 bases at least Q40

Consensus quality: 162109 bases at least Q30

Consensus quality: 164133 bases at least Q20

Estimated insert size: 161675; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 16506: contig of 16506 bp in length

\* 16507 16506: gap of unknown length

\* 180638: contig of 164032 bp in length

\* 180639 180738: gap of unknown length

\* 180739 182447: contig of 1709 bp in length

\* 182448 182547: gap of unknown length

\* 182548 195184: contig of 12637 bp in length.

## FEATURES

## source

1..195184

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-521F8"

466..1181

/notes="clone\_boundary"

/clone\_end:T7

site:

end\_sequence:BZ266854"

1768..2639

/note="clone boundary"

clone\_end:Sp6

site:

end\_sequence:BZ267695"

16807..17640

/note="wgs\_end\_extension"

clone\_end:Sp6"

71528..72576

/note="wgs\_end\_extension"

clone\_end:Sp6"

76693..78955

/note="wgs\_end\_extension"

clone\_end:Sp6"

86077..87329

/note="wgs\_end\_extension"

clone\_end:Sp6"

106905..108354

/note="wgs\_end\_extension"

clone\_end:Sp6"

129622..130958

/note="wgs\_end\_extension"

clone\_end:Sp6"

## ORIGIN

Query Match 64.0%; Score 773.4; DB 2; Length 195184;

Best Local Similarity 80.2%; Pred. No. 7e-75;

Matches 977; Conservative 0; Mismatches 217; Indels 24; Gaps 5;

QY 1 ATGAAGTTGAGAGTGTTCCTCCCTCGCGCGCCACCGGACACAGGAGGAGGAGTACTG 60

Db 136153 ATGAATTTGAGAGTATTTCCTCCCTCGCGCGCCACCGGACACAGGAGGAGTACTG 136212

QY 61 GAGGCG 120

Db 136213 GAGGCG 136272

QY 121 GGCTCAGATGGGACTCG 177

Db 136273 GGCTCAGCGGGACTGTGCG 136332

QY 178 CAGGCG 237

Db 136333 GAAGCG 136380

QY 238 GCAGTGTGTGAGCGGT 297

Db 136381 GACGATCCGAGGTGTGACCGATGGCAGCAGAACGCGAGGCGCTCCCGCGTGGCGCGCGCG 136440

QY 298 GCG 357

Db 136441 GCGAGCGTGGCGCGCGGT 136500

QY 358 CCCTACTCGTACATCGCGCTCATCGCCATCGCCATCGCCATCGCCATCGCCATCGCCATCG 417

Db 136501 CCCTACTCGTACATCGCGCTCATCGCCATCGCCATCGCCATCGCCATCGCCATCG 136560

QY 418 AGCTTGGCGGAGATCAACGAGTACTCATGTGGGGAAGTTCCTCTTTCCCGGGGAGCTAC 477

Db 136561 AGCTTGGCGGAGATCAACGAGTACTCATGTGGGGAAGTTCCTCTTTCCCGGGGAGCTAC 136620

```

Center: Code: wisk
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L10103
Center clone name: 403.I.20
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.940731
Consensus quality: 17335 bases at least Q40
Consensus quality: 177417 bases at least Q30
Consensus quality: 179105 bases at least Q20
Insert size: 1790000; agarose-fp
Insert size: 180452; sum-of-contigs
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs
-----
* NOTE: this is a 'working draft' sequence. It currently
  * consists of 18 contigs. The true order of the pieces

```

\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 2222: contig of 2222 bp in length
* 2222: gap of 100 bp
* 2223: contig of 1463 bp in length
* 2223: gap of 100 bp
* 3785: gap of 100 bp
* 3786: contig of 2868 bp in length
* 3786: gap of 100 bp
* 6754: gap of 100 bp
* 6754: contig of 3183 bp in length
* 6854: gap of 100 bp
* 10037: gap of 100 bp
* 10137: contig of 4401 bp in length
* 14538: gap of 100 bp
* 14638: contig of 6540 bp in length
* 21177: gap of 100 bp
* 21277: contig of 8963 bp in length
* 30241: gap of 100 bp
* 30241: contig of 9783 bp in length
* 40123: gap of 100 bp
* 40123: contig of 7776 bp in length
* 40224: gap of 100 bp
* 40224: contig of 10136 bp in length
* 48000: gap of 100 bp
* 48000: contig of 10136 bp in length
* 58236: gap of 100 bp
* 58236: contig of 8706 bp in length
* 67042: gap of 100 bp
* 67142: contig of 8780 bp in length
* 75922: gap of 100 bp
* 75922: contig of 10506 bp in length
* 86528: gap of 100 bp
* 86528: contig of 14707 bp in length
* 101335: gap of 100 bp
* 101335: contig of 16196 bp in length
* 117631: gap of 100 bp
* 117631: contig of 20563 bp in length
* 138294: gap of 100 bp
* 138294: contig of 19448 bp in length
* 157842: gap of 100 bp
* 157842: contig of 24211 bp in length.
  
```

## FEATURES

source

```

1. 182152
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="6"
  /map="6"
  /clone="RP11-403I20"
  /clone_lib="RPCI-11 Human Male BAC"
1. 2222
  /note="assembly_fragment"
  2323..3785
  /note="assembly_fragment"
  3886..6753
  /note="assembly_fragment"
  6854..10036
  /note="assembly_fragment"
  10137..14537
  /note="assembly_fragment"
  14638..21177
  /note="assembly_fragment"
  21278..30240
  /note="assembly_fragment"
  30341..40123
  /note="assembly_fragment"
  40224..47999
  /note="assembly_fragment"
  clone_end:SP6
  vector_side:left"
  48100..58235
  /note="assembly_fragment"
  58336..67041
  
```

```

/note="assembly_fragment"
67142..75921
/note="assembly_fragment"
76022..86527
/note="assembly_fragment"
86628..101334
/note="assembly_fragment"
101435..117630
/note="assembly_fragment"
117731..138293
/note="assembly_fragment"
138394..157841
/note="assembly_fragment"
157942..182152
/note="assembly_fragment"
clone_end:SP6
vector_side:right"

ORIGIN
Query Match      63.6%; Score 768.4; DB 2; Length 182152;
Best Local Similarity 87.8%; Pred. No. 2.5e-74;
Matches 772; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 331 AAGCATATACGGGGGCGCCAAAGCCCCCTACTGTACATCGGCTCATGCCATGCC 390
Db 10141 AACCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 10082
QY 391 ATCCGGGACTCGGGGGCGGGCTTGACGCTGGCGGAGATCAACGAGTAGTACCTCATGGC 450
Db 10081 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 10022
QY 451 AAGTTCCTCCCTTTTCCGCGGAGCTACACGGGCTGGCGCAACTCCGTCGCGCAACCTT 510
Db 10021 AAGTTCCTCCCTTTTCCGCGGAGCTACACGGGCTGGCGCAACTCCGTCGCGCAACCTT 962
QY 511 TCGCTCAACGACTGCTTCTGTCAGGTCGTCGGGAGCCCTCGCGGCGCTGGGGCAAGGAC 570
Db 9961 TCGCTCAACGACTGCTTCTGTCAGGTCGTCGGGAGCCCTCGCGGCGCTGGGGCAAGGAC 9902
QY 571 AACTACTGGATGCTCAACCCCAACAGCGAGTACACCTTCGCGGACGCGGCTTTCGCGCGC 630
Db 9901 AACTACTGGATGCTCAACCCCAACAGCGAGTACACCTTCGCGGACGCGGCTTTCGCGCGC 9842
QY 631 CGCGCAAGCGCTCAGCCACCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGAG 690
Db 9841 CGCGCAAGCGCTCAGCCACCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGAG 9782
QY 691 GCGCGGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750
Db 9781 GCGCGGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9722
QY 751 CGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 810
Db 9721 CGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 9662
QY 811 GCCATCGACAGCATCTTGGCAAGCCCTTCCGAGAGCGCGCTCAGGGACACGCGCCCC 870
Db 9661 GCCATCGACAGCATCTTGGCAAGCCCTTCCGAGAGCGCGCTCAGGGACACGCGCCCC 9602
QY 871 GGGACGACGCTTTCAGTGGGGCGCGCGCGCTTCCCGCGCGCTTCCCGCGCGCTC 930
Db 9601 GGGACGACGCTTTCAGTGGGGCGCGCGCGCTTCCCGCGCGCTTCCCGCGCGCTC 9542
QY 931 CTCGCGCGCGCGCGCTTGCAGGGCGCTTGTGCGCGCTTGTGCGGTACGCGGCGCGAGCCG 990
Db 9541 CTCGCGCGCGCGCGCTTGCAGGGCGCTTGTGCGCGCTTGTGCGGTACGCGGCGCGAGCCG 9482
QY 991 CGCGGCTTGGCGCGCGCGAGCGCGAGGTGCAACGACCGCGCGCGCGCTTCTCTGCTTGA 1050
Db 9481 CGCGGCTTGGCGCGCGCGAGCGCGAGGTGCAACGACCGCGCGCGCGCTTCTCTGCTTGA 9422
QY 1051 CCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1110
  
```

Db 9421 CCTCTCCCGCGCGCGCCGCCAAGCCACTCCGAGGCGCGCGCGCGCGCGCAC 9362

Qy 1111 CTGTACTGCTCCCTGCGGCTGCGCGAGCCCTGCGAGGCGGCTTAGTCCGNCGTCTCTGGC 1170

Db 9361 CTGTACTGCTCCCTGCGGCTGCGCGAGCCCTGCGAGGCGGCTCACTCGCGCGCTGGC 9302

Qy 1171 CGGCACCTGCTGTAACCGGTGAGACGCTTCCTAGCTTGA 1209

Db 9301 CGGCACCTGCTGTAACCGGTGAGACGCTTCCTAGCTTGA 9263

RESULT 12

RATHFH1

LOCUS Rattus norvegicus HNF-3/forkhead homolog-1 (HNF-1) mRNA, complete cds.

DEFINITION Rattus norvegicus HNF-3/forkhead homolog-1 (HNF-1) mRNA, complete cds.

ACCESSION L13201.1 GI:951067

VERSION

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1760)

AUTHORS Clevidence D.E., Overdier, D.G., Tao, W., Qian, X., Pani, L., Lai, E. and Costa, R.H.

TITLE Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (9), 3948-3952 (1993)

MEDLINE 93248207

PUBMED 7683413

COMMENT On Aug 22, 1995 this sequence version replaced gi:550512.

Original source text: rat.

FEATURES

source

1..1760

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/sex="male"

/tissue\_type="lung"

/dev\_stage="9 week old"

/tissue\_lib="Clontech"

1..1760

/gene="HNF-1"

250..1569

/gene="HNF-1"

/notes="bp 84...889 winged-helix DNA binding domain"

/codon\_start=1

/product="HNF-3/forkhead homolog-1"

/protein\_id="AAA74561.1"

/db\_xref="GI:550513"

/translation="MGSDLGAGSDVPSPLSAAGDLSLGSDDCAANSPPAAGSAGD  
LEGGERNSSGASQDDPVEYDGSRTQASVPGCAGSVGGGKSKPVTTRPKPP  
YSTIALIAIRDSAGRLTALIEYELMGKFFPFGSVTGRNVSRLNLSINDCFVK  
VLEDPSPKQNYMWNENSETTFADGVRRRLKLSHRTVTSAGSGGSGPPGAG  
TQOPATGSSPIARSPPAREGSSPASKSFADISLSPFRSRDGLRWGCSY  
PGALLPARCAPTPRPPRPVPCRSVLTARRAHAGVARGAARGAPVRGAPLH  
RGPSQALSRGRRRGAAPVLPPTAAHPAGRGRLRSAPVLPDGGDAASLTGAWSSAG  
GSRVELGAPVMEGTCHGWNAGSAVHPSPRNRDFFSTLHQMC"

gene

CDS

Query Match 59.4%; Score 717.8; DB 10; Length 1760;

Best Local Similarity 78.7%; Pred. No. 3.1e-68;

Matches 951; Conservative 0; Mismatches 233; Indels 24; Gaps 7;

Qy 8 TGGAGTGTTCGTCCTCGCGCGCGCCCGGACAGCAGGCGGACGTGACCTGGAGGGG 67

Db 213 TTGGAGTGTTCGTCCTCGCGCGCGCCCGGACAGCAGGCGGACGTGACCTGGAGGGG 271

Qy 68 CGGCGCGCAGCAGCGGCGCGCTCCCGCTGTGCGCGCGCGGAGACGACTCCCTGGGCTCAG 127

Db 272 CCGGCGAGCAGCAGCTGGCCATCTCCACTGTCCCGGCGCTGGTACGACTCCTTAGGCTCAG 331

Qy 128 ATGGGGACTGCGCGGCCAA---GCCGTCCGCGCGCGCGCGCCGACAGATACGAGAGGCG 184

Db 332 ACGGGGACTGTGCAGCCAAACAGCCGCGCGCGCGCAGCGCGCGGGATCTGGAAGGTG 391

Qy 185 ACGGCGAAGAGTGGGAGAGGCGCGCGCGCGCGGAGGAGCGATCCCGCGCAGCAGCTG 244

Db 392 CGCGCGCGAGAGGAATTCAGTGG-----CGGGCGAGCACCCAGACGATC 439

Qy 245 CTGCAGCGTGTGGCGAGGCGCGCGAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 304

Db 440 CCGAGGTGACCGATGGCAGCAGAAACGAGGCTCCCGCGTGGGGCGCTGTGCGCGGCGAG 499

Qy 305 CGGGAGCGCGGAGGCTGCACGAGCAGCAAGCATATACGCGCGCGCGCGCGCGCGCGCTACT 364

Db 500 TGGCGCGCGGTGAGGCGCGCGCAGCAAGCCGTACACGCGCGCGCGCGCGCGCGCGCTACT 559

Qy 365 CGTACATCGCGCTCATCGCCATCGCCATCGCGACTCGCGCGCGCGCGCGCTTGAAGCTGG 424

Db 560 CTTACATCGCACTCATCGCCATCGCCATCGCGACTCGCGCGCGCGCGCTGACGCTGG 619

Qy 425 CGGAGATCAAGAGTACTCATGGGGAAGTTCCTCTTCGCGCGCGCGCGCGCGCGCGCTACT 484

Db 620 CCGAGATCAAGGATCTCATGGGCAAGTTCCTCTTCGCGCGCGCGCGCGCGCGCGCTACT 679

Qy 485 GGCGCAACTCGCTGCGCGCACAACTTTCGCTCAACGACTGCTTCGTCGTCGTCGTCGCG 544

Db 580 GGCGCACTCGCTGCGCGCACAACTTTCGCTCAACGACTGCTTCGTCGTCGTCGTCGCG 739

Qy 545 ACCCTCGCGCGCTCGCGCGCAAGCAACTACTTGGATGCTCAACCCCAACAGCGAGTACA 604

Db 740 ACCCTCGCGCGCTCGCGCGCAAGCAACTACTTGGATGCTCAACCCCAACAGCGAATACA 799

Qy 605 CTTTCGCGCGCGCGCTTCGCGCGCGCGCGCAAGCGCTCAGCACCGCGCGCGCGCTCC 664

Db 800 CTTTCGCGCGCGCGCTTCGCGCGCGCGCGCAAGCGCTCAGCACCGCGCGCGCTCC 859

Qy 665 CCGCGCGCGCGCTGCG 724

Db 860 CCGCATCGGCTACCGCGGAGGAAGCCACCGCGGAGCGCGCGCGCGCGCGCGCGCG 919

Qy 725 CCGCGCGCGCGCGCTCG 784

Db 920 CCACCGCGCGCGCTCTCCCGCAATCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 979

Qy 785 CCGCGCGCGCGCTCTCCAGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 844

Db 980 CCGCGAGCAAGTCTCTCCAGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1039

Qy 845 GCCGTGCGCTCAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 904

Db 1040 GCCCGCGCGAC-GGCACCGCGCTCTGGGGGTGCGAGTACCGTGGAGCGCTGCTCCCTGCC 1098

Qy 905 CGCGCTGCGCGCGCTTCGCGCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 964

Db 1099 CGCGCTGCGCGCGCTATCCCGCGCTCTCTCCCGCGCTCGTCCCGCGGTGCGCTGTCGCG 1158

Qy 965 TCTCGCGCTACG 1024

Db 1159 TCTGTGTTTACGGCGC-GGCGAGCCACGCTGTGCGGTGCGCGCGCGCGCGCGCGCG 1214

Qy 1025 CGACCGCGCGCGCGCTCTGCTGTCACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1084

Db 1215 AGCCCGCGCGCGCGCTGTTCGTGGCGCGCGCTCTCCACCGCGCGCGCGCGCGCGCG 1274

Qy 1095 GAGCGCGCG---CG 1141

Db 1275 GAGTCCCGGAGACCG 1394

Qy 1142 TCGAGCGCGCGCTTAGTCCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1201

Db 1335 TGCAGGCGCGCGCGCGCGCTGCGGTCCGCGTCCGCGACCTGTCTTACCGGGTGGAGCGCTGC 1394



Unpublished

2 (bases 1 to 74969)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B., Brown, A., Castle, A., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Doming, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lechoczky, J., Lieuw, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:7144942.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: W18R  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L3560  
Center clone name: 16 J 9

```

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

1	944:	contig of 944 bp in length
*	1044:	gap of 100 bp
*	945	1044:
*	1045	contig of 947 bp in length
*	1991:	contig of 100 bp
*	1992	2091:
*	2091:	gap of 916 bp in length
*	2092	3007:
*	3008	3107:
*	3108	gap of 100 bp
*	4037:	contig of 930 bp in length
*	4137:	gap of 100 bp
*	4138	4137:
*	5056:	contig of 919 bp in length
*	5057	5156:
*	5156:	gap of 100 bp
*	6079:	contig of 923 bp in length
*	6179:	gap of 100 bp
*	6080	6179:
*	6180	contig of 924 bp in length
*	7104	7203:
*	7204	8136:
*	8137	contig of 933 bp in length
*	8237	8236:
*	9166:	gap of 100 bp
*	9167	9166:
*	9267	contig of 930 bp in length
*	10167:	gap of 100 bp
*	10168	10267:
*	10268	gap of 100 bp
*	11212:	contig of 945 bp in length
*	11213:	contig of 100 bp
*	11313	11312:
*	12250:	contig of 938 bp in length
*	12251	12350:
*	12351	gap of 100 bp
*	13277:	contig of 927 bp in length
*	13278	13377:
*	13378	gap of 100 bp
*	14275:	contig of 898 bp in length
*	14375:	gap of 100 bp
*	14376	14375:
*	15381:	contig of 906 bp in length
*	15382	15381:
*		gap of 100 bp

RESULT 14	
LOCUS	AC016269/c
LOCUS	74969 bp DNA linear HTG 13-JUL-2000
DEFINITION	Homo sapiens clone RP11-16J9, LOW-PASS SEQUENCE SAMPLING.
DEFINITION	AC016269
VERSION	AC016269.3 GI:9123861
KEYWORDS	HTG; HTGS PHASE0.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 74969)
TITLE	Biren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE	Homo sapiens clone RP11-16J9



```

Db      38291  CCTACTCGTACATCGCGCTCATCGCCATCGCATCCGAGCTCGGCGGGCGGCTTG 38232
Qy      418  ACCTGCGGAGATCAACAGTACTCTCATGCGGAGTTCCTCCCTTTTTCGCGGCGAGCTAC 477
Db      38231  ACCTGCGGAGATCAACAGTACTCTCATGCGGAGTTCCTCCCTTTTTCGCGGCGAGCTAC 38172
Qy      478  ACGGGCTCGGCAACTCCGTCGCGCCACACCTTTTCCTCAACGACTGCTTCGTCAAGGTG 537
Db      38171  ACGGGCTCGGCAACTCCGTCGCGCCACACCTTTTCCTCAACGACTGCTTCGTCAAGGTG 38112
Qy      538  CTGCGGCACTCTCGGCGGCTCGGCGCAAGGACAACTACTGATGCTCAACCCCAACAGC 597
Db      38111  CTGCGGCACTCTCGGCGGCTCGGCGCAAGGACAACTACTGATGCTCAACCCCAACAGC 38053
Qy      598  GAGTACACCTTCGCCGACG 616
Db      38052  GGGGGGGGCTTAGAGTCG 38034

RESULT 15
AC067929
LOCUS   182152 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 6 clone RP11-403120 map 6, WORKING DRAFT
SEQUENCE, 18 unordered pieces.
AC067929
VERSION AC067929.2 GI:8247824
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 182152)
Birren, B., Linton, L., Nuebaum, C. and Lander, B.
Homo sapiens chromosome 6, clone RP11-403120
Unpublished
2 (bases 1 to 182152)
Birren, B., Linton, L., Nuebaum, C., Lander, B., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepe, Y., Colangelo, M., Collins, S.,
Collimore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, F., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182152)
Birren, B., Linton, L., Nuebaum, C., Lander, B., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepe, Y., Colangelo, M., Collins, S.,
Collimore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,

```

```

Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, F., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 4, 2000 this sequence version replaced gi:7655991.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10103
Center clone name: 403 I-20
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173395 bases at least Q40
Consensus quality: 177417 bases at least Q30
Consensus quality: 179105 bases at least Q20
Insert size: 179000; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 2222: contig of 2222 bp in length
* 2322: gap of 100 bp
* 2323
* 3785: contig of 1463 bp in length
* 3786
* 3883: gap of 100 bp
* 3886
* 6753: contig of 2868 bp in length
* 6754
* 6853: gap of 100 bp
* 6854
* 10036: contig of 3183 bp in length
* 10037
* 10136: gap of 100 bp
* 10137
* 14537: contig of 4401 bp in length
* 14538
* 14637: gap of 100 bp
* 14638
* 21177: contig of 6540 bp in length
* 21178
* 30240: contig of 8963 bp in length
* 30241
* 30340: gap of 100 bp
* 30341
* 40123: contig of 9783 bp in length
* 40124
* 40223: gap of 100 bp
* 40224
* 47998: contig of 7776 bp in length
* 48000
* 48099: gap of 100 bp
* 48100
* 58235: contig of 10136 bp in length
* 58236
* 58335: gap of 100 bp
* 58336
* 67041: contig of 8706 bp in length
* 67141: gap of 100 bp
* 67142
* 75921: contig of 8780 bp in length
* 75922
* 76021: gap of 100 bp
* 76022
* 86527: contig of 10506 bp in length
* 86528
* 101334: contig of 14707 bp in length
* 86628
* 101334: contig of 14707 bp in length
* 101335
* 101434: gap of 100 bp
* 101435
* 117630: contig of 16196 bp in length

```

TITLE  
JOURNAL  
COMMENT

*	117631	117730: gap of 100 bp
*	117731	138293: contig of 20563 bp in length
*	138294	138393: gap of 100 bp
*	138394	157841: contig of 19448 bp in length
*	157842	157941: gap of 100 bp
*	157942	182152: contig of 24211 bp in length.

## FEATURES

```

source
1. 182152
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-403120"
/clone_lib="RPC1-11 Human Male BAC"

1. 2232
/note="assembly_fragment"
2323. 3785
/note="assembly_fragment"
3886. 6753
/note="assembly_fragment"
6854. 10036
/note="assembly_fragment"
10137. 14537
/note="assembly_fragment"
14638. 21177
/note="assembly_fragment"
21278. 30240
/note="assembly_fragment"
30341. 40123
/note="assembly_fragment"
40224. 47999
/note="assembly_fragment
vector_side:left"
48100. 58235
/note="assembly_fragment"
58336. 67041
/note="assembly_fragment"
67142. 75921
/note="assembly_fragment"
76022. 86527
/note="assembly_fragment"
86628. 101334
/note="assembly_fragment"
101435. 117630
/note="assembly_fragment"
117731. 138293
/note="assembly_fragment"
138394. 157841
/note="assembly_fragment"
157942. 182152
/note="assembly_fragment
vector_side:right"

```

## ORIGIN

	Query Match	20.7%	Score 250.8;	DB 2;	Length 182152;
	Best Local Similarity	71.6%;	Pred. No. 6.1e-19;		
	Matches 265;	Conservative 0;	Mismatches 102;	Indels 3;	Gaps 1;
Qy	1	ATGAAGTTGAGGTGTTCTGTCCTCTCGGGCGGCCACAGCGGGACAAGCAGGGCAGTGACCTG	60		
Db	29973	ATGAAGTTGAGGTGTTCTGTCCTCTCGGGCGGCCACAGCGGGACAAGCAGGGCAGTGACCTG	30032		
Qy	61	GAGGGCGGGCGGCAGCGACGCGTGTCCCGCTGTCTGGCGCGCGGAGACCACTCCCTG	120		
Db	30033	GAGGGCGGGCGGCAGCGACGCGTGTCCCGCTGTCTGGCGCGGAGACCACTCCCTG	30092		
Qy	121	GGCTCAGATCGGGAAGTCTCGCGGCCAA---GCCTTCGCGGGCGGGGCCACAGATACG	177		
b	30093	GGCTCAGATCGGGAATCTCGCGGCCAAACAGCGCCGGCGCGCGGGCGGGCGGCCACAGATACG	30152		

Qy	178	CAGGGCGACGGCGAAACAGAGTCCGGAGGCGGGCGCGCGAGGAGGCATCCCCGCA	237
Db	30153	CAGGGCGACGGCGAAACAGAGTCCGGAGGCGGGCGCGCGAGGAGGCATCCCCGCA	30212
Qy	238	GCAGCTGCTGCAGCGGTGTCGAGAGGCGCGAGGCGCGGGCGCGGAGCGCG	297
Db	30213	GCAGCTGCTGCAGCGGTGTCGAGAGGCGCGAGGCGCGGGCGCGGAGCGCG	30272
Qy	298	GGCGGCGCGGGAGCGCGAGGGTGCACGACAGCCATATACGCGCGCGCCCAAGCCC	357
Db	30273	NN	30332
Qy	358	CCCTACTCGT	367
Db	30333	NNNNNNNGT	30342

Search completed: March 31, 2004, 00:21:43  
Job time : 4698 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 21:49:37 ; Search time 502 Seconds

(without alignments)  
10231.218 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209  
Sequence: 1 atgaagtggagggttcgt.....tggagacgctcctagcttga 1209

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002s:\*
- 7: Geneseqn2003as:\*
- 8: Geneseqn2003bs:\*
- 9: Geneseqn2003cs:\*
- 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1208	99.9	1209	6	ABQ81557 Gene up-r
2	717.8	59.4	1760	7	ABT42112 Toxicity
3	442	36.6	658	6	ABQ39023 Oligonucle
C	442	36.6	658	6	ABQ39022 Oligonucle
C	305.4	25.3	658	6	ABQ39025 Oligonucle
6	305.4	25.3	658	6	ABQ39024 Oligonucle
7	201.4	16.7	2487	9	ABD76311 Prostate
8	201.4	16.7	5181	6	ABL68935 Kidney ca
9	195	16.1	1380	9	ADD29558 Human tum
10	194.6	16.1	3482	7	ABX76288 Lung canc
11	193.4	16.0	2271	6	ABZ35021 Human gen
12	187.6	15.5	1860	6	AAQ50632 Brain fac
13	184.4	15.3	1965	9	ADB31305 Testoster
14	184.4	15.3	2187	7	ABZ34783 Coding se
15	175.6	14.5	3946	2	AZ33671 Human FK
16	175.6	14.5	3946	2	AZ33671 Human FK
17	175.2	14.5	2512	5	AAZ67345 DNA encod
18	171.6	14.2	1662	2	AAZ31672 Human FK
19	171.6	14.2	1662	2	AZ338080 Human for
20	165.6	13.7	1976	2	AZ28095 Human FRE
21	161.4	13.3	2745	7	AZ220402 Oncofeta
22	161.4	13.3	2753	6	ABT03399 Ovary cel
23	160.4	13.3	1355	7	ABZ34708 Coding se

24	153.8	12.7	2934	6	AB199306	Abi99306 Mouse isc
25	153.8	12.7	3289	4	AAH26570	Aah26570 Human win
26	153.8	12.7	6458	6	ABK48984	Abk48984 Genomic D
27	152.2	12.6	2106	2	AAK28103	Aax28103 Freac11 g
28	151	12.5	3786	5	AAK2683	Aas2683 DNA encod
29	150.4	12.4	2448	9	ADD29815	Add29815 Human tum
30	149.6	12.4	2832	2	AAQ50631	Aaq50631 Brain fac
31	146.2	12.1	1287	5	AAK2684	Aas2684 DNA encod
32	143.4	11.9	2559	9	ADD15216	Add15216 Human ser
33	140.4	11.6	1422	6	ABK15142	Abk15142 Human hep
34	140.4	11.6	2872	6	ABL69589	Ab169589 Prostate
35	140.4	11.6	2872	6	ABT11015	Abt11015 Human bre
36	140.4	11.6	2872	7	ACC50184	Acc50184 Breast ca
37	140.4	11.6	3098	5	ABV25148	Abv25148 Human pro
38	140.4	11.6	11613	6	ABK15141	Abk15141 Human hep
C	139	11.5	1711	6	ABQ50740	Abq50740 Oligonucle
40	139	11.5	1711	6	ABQ50741	Abq50741 Oligonucle
41	136.2	11.3	812	6	ABQ61118	Abq61118 Transcrip
42	136	11.2	441	7	ABX41089	Abx41089 Bovine ES
43	134.8	11.1	4450	4	AAK90986	Aak90986 Human dig
44	134.8	11.1	4450	5	AAK32021	Aas32021 Human liv
45	134.8	11.1	4450	6	ABN90376	Abn90376 Human liv

#### ALIGNMENTS

##### RESULT 1

ABQ81557  
ID ABQ81557 standard; cDNA; 1209 BP.

XX  
AC ABQ81557;

XX  
DT 30-DSC-2002 (first entry)

XX  
DE Gene up-regulated in metastatic colorectal cancer.

XX  
KW Colorectal cancer; metastasis; differential expression; cytostatic;

XX  
KW diagnosis; gene therapy; vaccine; gene; ss.

XX  
OS Homo sapiens.

XX  
PH Key Location/Qualifiers

FT CDS 1..1209

FT /\*tag= a

FT /product= "Metastatic colon cancer polypeptide"

XX  
PN WO200268677-A2.

XX  
PD 06-SEP-2002.

XX  
PF 27-FEB-2002; 2002WO-US006001.

XX  
PR 27-FEB-2001; 2001US-0272206P.

XX  
PR 02-APR-2001; 2001US-0281149P.

XX  
PR 17-APR-2001; 2001US-0284555P.

XX  
PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX  
PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX  
XX Mack DH, Markowitz SD;

XX  
XX WPI; 2002-698677/75.

XX  
XX P-PSDB; ABP54631.

XX  
XX New genes that are up- or down-regulated in colorectal cancer, useful for  
diagnosing colorectal cancer in a subject, or for identifying modulators  
of colorectal cancer-associated proteins and genes for treating  
colorectal cancer.

XX  
PS Claim 5; Page 252; 260pp; English.

XX  
CC The present sequence is the nucleotide sequence of a human gene that



XX (GENE-) GENE LOGIC INC.  
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;  
XX WPI; 2003-148464/14.  
XX  
XX Predicting at least one toxic effect of a compound, useful for toxicity  
PT modeling, comprises preparing a gene expression profile of a tissue or  
PT cell sample exposed to the compound, and comparing the gene expression  
PT profile to a database.  
XX  
XX Example 4; Page; 446pp; English.  
XX  
XX The invention relates to a novel method of predicting at least one toxic  
CC effect of a compound. The method comprises a gene expression profile of a  
CC tissue or cell sample exposed to the compound, and comparing the gene  
CC expression profile to a database comprising at least part of the data or  
CC information given in the specification. The methods are useful for  
CC predicting at least one toxic effect of a compound, predicting the  
CC progression of a toxic effect of a compound, predicting the renal  
CC toxicity of a compound, or identifying toxicity markers in tissues or  
CC cells exposed to known renal toxin. The genes are useful as toxicity  
CC markers in drug screening and toxicity assays, in monitoring disease or  
CC physiological states, or disease progression. This polynucleotide  
CC represents a rat DNA sequence relating to the toxic effect database  
CC described in the specification. NOTE: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from the World Intellectual Property  
CC Organization  
XX  
SQ Sequence 1760 BP; 301 A; 615 C; 535 G; 309 T; 0 U; 0 Other;  
Query Match 59.4%; Score 717.8; DB 7; Length 1760;  
Best Local Similarity 78.7%; Pred. No. 3.2e-87;  
Matches 951; Conservative 0; Mismatches 233; Indels 24; Gaps 7;  
8 TTGAGGTGTTGCTCTCCGCGGCGCCACGCGGCAAGCAGGCGATGACCTGGAGGGCG 67  
213 TTGAGGTGTTGCTCTCCGCGGCGCCACGCGGCAAGCAGGCGATGACCTGGAGGGCG 271  
68 CGGCGCGGCGGCGGCGGCTCCCGCTCTCGGCGCGGCGGAGACGACTCCCTGGGCTCAG 127  
272 CCGGCGGCGGCGGCGGCTCTCCGCTCTCGGCGGCGGAGTGTGACGACTCTTAGGCTCAG 331  
128 ATGGGAGCTGCGGCGGCGCAA---GCCGTCCGCGGCGGCGGCGCCAGAGATACGCGAGGCG 184  
332 ACGGGAGCTGTGCGAGCAACAGCGCGCGGCGGCGGCGCGCGGATCTGGAAGGTG 391  
185 ACGCGCAACAGAGTGGGAGGCGGCGCGGCGGCGGCGGAGGCGGATCCCGGCGAGCTG 244  
392 GCGCGCGGCGGAGGGAATTCGAGTGG-----CGGGCGGAGCACCCGAAGACGATC 439  
245 CTGCGAGCGGTGTTGGCGGAGGCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 304  
440 CCGAGTGACCGATGCGCAGCAGCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 499  
305 CGGGGAGCGCGGAGGTGTACGCGAGCAAGCATATAGCGGCGGCGGCGGCGGCGGCGGCTACT 364  
500 TGGCGCGCGGTGAGGGCGGCGGCGGAGCAAGCGGTACACGCGGCGGCGGCGGCGGCGGCTACT 559  
365 CGTACATCGGCGCTCATCGGCATCGGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 424  
560 CTTACATCGCACTCATCGGCATCGGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 619  
425 CGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTTTCGCGGCGAGCTACACGGGCT 484  
620 CCGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTTTCGCGGCGAGCTACACGGGCT 679  
485 GGGCGCACTCCGCGGCGCAACACTTCGCTCAACGAGCTGCTTCGTCAGGCTGTCGCGG 544  
680 GGGCGCACTCCGCGGCGCAACACTTCGCTCAACGAGCTGCTTCGTCAGGCTGTCGCGG 739

QY 545 ACCCTCGCGGCGCTTGGGGCAAGACAACTACTGGATGCTCAACCCCAACAGCGAGTACA 604  
DB 740 ACCCTCGCGGCGCTTGGGGCAAGACAACTACTGGATGCTCAACCCCAACAGCGAGTACA 799  
QY 605 CTTTCGCGGCGAGGGGTCTTCCGCGCGCGCGCAAGCGCTTCAGCCACCGCGCGCGCGTCC 664  
DB 800 CTTTCGCGGCGAGGGGTCTTCCGCGCGCGCGCGCAAGCGCTTCAGCCACCGCGCGCGT 859  
QY 665 CCGCGCGCGGGTCTCGCGCGCGAGAGGCGCGGGCGCTCCCGCGCGCGCGCGCGCGCG 724  
DB 860 CCGCATCGGGCTACGCGCGGAGGAGACCCCGCGGACCTGCGGGGACCCCGCGCGCGCG 919  
QY 725 CCGCGCGCGCGCGCTCGCGCGCGATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 784  
DB 920 CCACCGCGGCTCTCTCCCAATCGCGCGCTCGCGCGCTCGCGAGGAGGGCTCCAGCG 979  
QY 785 CCGCGCGGAGAGTTCTCAGCTCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 844  
DB 980 CCGCGGAGCAAGTTCTCAGCTCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1039  
QY 845 GCCGTGCGCTCAGGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 904  
DB 1040 GCCCGCGGAG-GGACCCCGGCTCTGGGGGTGAGCTACCTGGAGGCTGCTCCCTGCC 1098  
QY 905 CGCGCTGCGCGGCTTCCCGCGGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 964  
DB 1099 CGCGCTGCGCGGCTTCCCGCGGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1158  
QY 965 TCTCGCGGTACGGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1024  
DB 1159 TCTGTGTTACGGCGG-GGCGAGCCAGCTGCTGGCGTGGCGGCGCGCGCGCGG 1214  
QY 1025 CGACCGCGCGCGCGCTCTCTGTTGCACTCTCCCGCGGCGGCGCGCGCGCGCGCG 1084  
DB 1215 AGCGCGCGGCGCGCTGTTGTTGCGCGCGCTCTCCACCGCGGCGCGCGCGCGCG 1274  
QY 1085 GAGCGCGCGGCG 1141  
DB 1275 GAGTTCGCGGAGACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1334  
QY 1142 TGACGCGGCGCTTAGTCCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1201  
DB 1335 TGCAGGCG 1394  
QY 1202 TAGCTTGA 1209  
DB 1395 TAGCTTGA 1402  
RESULT 3  
ABQ39023  
ID ABQ39023 standard; DNA; 658 BP.  
XX ABQ39023;  
XX 12-JUL-2002 (first entry)  
XX  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 25614.  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.  
XX Homo sapiens.  
XX WO200218632-A2.  
XX 07-MAR-2002.  
XX 01-SEP-2001; 2001WO-EP010074.  
XX 01-SEP-2000; 2000DE-01043826.  
PR

PR 05-SEP-2000; 2000DE-01044543.  
XX (EPiG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX  
XX Sequence 658 BP; 173 A; 310 C; 108 G; 67 T; 0 U; 0 Other;  
Query Match 36.6%; Score 442; DB 6; Length 658;  
Best Local Similarity 82.7%; Pred. No. 1.6e-50;  
Matches 505; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
599 AGTACACTTCGGCGAGGGGTCTTCGGCGCGCGCCGACAGCGCTCAGCACCGCGCG 658  
2 AATACACTTCGGCGAGGAATCTTCGGCGCGCGCCGACAGCGCTCAGCACCGCGCG 61  
659 CGGTCCCGCGCGCGCGCTCGCGCGCGAGAGGCGCGCGCGCTTCGGCGCGCGCGCG 718  
62 CGATCCCGCGCGCGCGAGTACGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 121  
719 CCGCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 778  
122 CCG 181  
779 CCAGCG 838  
182 CCAACCG 241  
839 TCCG 898  
242 TCCG 301  
899 CCTCG 958  
302 CCTACCG 361  
959 TCCG 1018  
362 TACCGAA 421  
1019 TGCCACCGAGC 1078  
422 TACCACCGAAC 481

QY 1079 CACTCCGAGGCG 1138  
DB 482 CACTCCGAGGCG 541  
QY 1139 CCCTGCGAGGCG 1198  
DB 542 CCCTGCGAGGCG 601  
QY 1199 TCCTAGCTTGA 1209  
DB 602 TCCTAACCTAA 612  
RESULT 4  
ID ABQ39022/c  
XX ABQ39022 standard; DNA; 658 BP.  
AC ABQ39022;  
XX 12-JUL-2002 (first entry)  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25613.  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
OS Homo sapiens.  
XX WO200218632-A2.  
XX 07-MAR-2002.  
XX 01-SEP-2001; 2001WO-EP010074.  
XX 01-SEP-2000; 2000DE-01043826.  
XX 05-SEP-2000; 2000DE-01044543.  
PA (EPiG-) EPIGENOMICS AG.  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX  
XX Sequence 658 BP; 67 A; 108 C; 310 G; 173 T; 0 U; 0 Other;

	Query Match	35.6%; Score 442; DB 6; Length 658;
	Best Local Similarity	82.7%; Pred. No. 1.6e-50;
	Matches 505; Conservative	0; Mismatches 106; Indels 0; Gaps 0;
Qy	599	AGTACACTTTGGCCGACGGGTCTTCGCGCCGCCCGCAAGCGCCTCAGCCACGCGCGGC 658
Dd	657	AATACACTTTGGCCGAGAAATCTTCGCGCGCGCGCAACGCTCAACCACGCGCGC 598
Qy	659	CGGTCCCCGCGCGCGGCTGCGGCCCGGAGGAGGCCGCGGCTCCCGCGCCCGCGCGC 718
Dd	597	CGATTCCC CGCGCGCGAATCTAGACCCGAATAAACCCCGAACTCCCGCGCGCCCGCGC 538
Qy	719	CCGCGCCCGCGCGCCCGCGCTTCGCCCCGCGATCGCTCGCCCGCGCGCACGAGGAGCGC 778
Dd	537	CCGCGCGCGCGCGCCGACCTTCGCCCCGCGATACGCTCGCGCGCGCGCAAAAAACGCG 478
Qy	779	CCAGCCCGCGGGCAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTGCGCAAGCCT 838
Dd	477	CCAAACCCCGGAAACAATCTCCAACTCTTCGCCATCGACCAATCTTAGCAAAACCT 418
Qy	839	TCCGACGCGCTCGCCTCAGGAGACACGCCCCCGGAGCAGCGCTTCAGTGCGGCGCGCGC 898
Dd	417	TCCGCAACCGCGGCTCAAAAACACGACCCCGGAAACGAGCTTCAATAAAACGCGCGC 358
Qy	899	CCTGCCCGCGCTGCCGCGCTTCCCGCGCTCCTCCCGGCGCTCCCGCGGCGCCTGCGAGCCCTGC 958
Dd	357	CCTACCCGCGCTACCCGCGTTCCCGCGCTTCCCGCGGACGCCCTACAAAACCTTAC 298
Qy	959	TGCGCTCTCGGCTACGCGCGCGGCGAGCGCGCGCTGGCGCGCGCGAGGCGCGAGG 1018
Dd	297	TACCGCTCTACGCTAGCAGCGAAACGAAACGACGCGACTAAACGCGCGGAAACCGAAA 238
Qy	1019	TGCCACGACCGGCGCGCCCTCTGCTTGCACTCTCCCGGCGCGCGCCCGCGCGCGCGC 1078
Dd	237	TACCACGACCGCGCGCCCTCTACTTACCTCTCCGACGACGACCCCGCGCAAC 178
Qy	1079	CACTTCGAGCGCGCGCGCGCGCGCGCGCTGTACTGTACTGCCCTCGGCTGCCCGCAG 1138
Dd	177	CACCTCCGAAACCGGACGACCGACGACGCGACCTATACTACCCCTACGACTACCCGCAA 118
Qy	1139	CCCTGACGGCGGCTTAGTCGNGCTCTGCGCGCACCTGTGCTACCCGCTGGAGACGC 1198
Dd	117	CCCTACAAAGACTCAATTCGCGCGCCCTTAACCGCGACCTACCGTACCCGATATAAACGC 58
Qy	1199	TCCTAGCTTGA 1209
Dd	57	TCCTAACCTAA 47
 RESULT 5 ABQ39025/c ID ID ABQ39025 standard; DNA; 658 BP.		
XX	AC	ABQ39025;
XX	DT	12-JUL-2002 (first entry)
XX	DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 25615.
KW	XW	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW	XW	drug; side effect; cancer; central nervous system; cardiovascular;
KW	XW	gastrointestinal; respiratory system; single nucleotide polymorphism;
OS	XX	SNP; cell differentiation; ds.
OS	XX	Homo sapiens.
DN	XX	WO200218632-A2.
PD	XX	07-MAR-2002.
PP	XX	01-SEP-2001; 2001WO-EP010074.
XX	XX	

PR	01-SEP-2000; 2000DE-01043826.
FR	05-SEP-2000; 2000DE-0104543.
XX	(EPIC-) EPIGENOMICS AG.
PA	Olek A, Piepenbrock C, Berlin K, Guetig D;
XX	WPI; 2002-371829/40.
DR	Determining the degree of cytosine methylation in genomic DNA, useful for
PT	diagnosis and prognosis, comprises selective hybridization of amplicons
PT	from chemically treated DNA.
XX	Claim 12; 56pp + Sequence Listing; 56pp; German.
PS	This invention describes a novel method for determining the degree of
XX	methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC	genomic sample of DNA. The sample is treated chemically to convert
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic
CC	DNA that contains the target C is amplified to form a labeled amplicon.
CC	The amplicon is hybridised to two classes, each with at least one member,
CC	of oligonucleotides and/or peptidenucleic acid (PNA) oligomers and the
CC	degree of hybridisation to both classes is determined from the label on
CC	the amplicon. From the ratio of labels hybridised to the two classes of
CC	oligomers, the degree of methylation is calculated. The method is used:
CC	(i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC	and of a wide range of diseases, e.g. cancer, disorders of the central
CC	nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC	particularly by detecting mutations or single nucleotide polymorphisms
CC	(SNPs); and (ii) for differentiation of cell or tissue types and for
CC	investigating cell differentiation. The method allows the methylation
CC	status of many C residues to be determined simultaneously. ABQ13410-
CC	ABQ54121 represent genomic DNA sequences used to illustrate the method
CC	for determining the degree of cytosine methylation described in the
CC	disclosure of the invention
XX	
SQ	Sequence 658 BP; 269 A; 222 C; 108 G; 59 T; 0 U; 0 Other;
Query Match            25.3%; Score 305.4; DB 6; Length 658;	
Best Local Similarity 68.6%; Pred. No. 2.le-32;	
Matches 420; Conservative         0; Mismatches 192; Indels      0; Gaps      0;	
QY	598 GAGTACACTTCGGCGACGGGGTCTCCGGCGCGCGAGAGCCCTCAGCACCGCGCG 657
Db	
Db	658 GAGTATATTTCGTGCACGGGGTTTTTCGTCTGCTGTGTAAGCGTTTAGTTATCGGCG 599
QY	658 CCGGTCCCCGCGCCCGGGCTGCGGCCCGAGAGAGCCCGGGCTCCCGCGCCCGCCGCG 717
Db	
Db	598 TCGGTTTTCGCTTGCGGTTTCGGTTCGAGAGGTTTCGGGTTTTTCGTCGTTTCGTGCG 539
QY	718 CCGCGCGCGCGCGCCCGGCTCCCGCGCATCGGCTCGCCCGCCCGCCGAGGAGCGCG 777
Db	
Db	538 TTCGCGTTCGTGTTTCGGTTTCGTTTCGTATCGGTTTCGTTCGTTTAGGAGGAGCGC 479
QY	778 GCACAGCCCCCGGGCAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTCTCGCAAGCCC 837
Db	
Db	478 GTTAGTTTCGGGGTAAATTTTTTTAGTTTTTCGTTATCGATAGTATTTTCGGTAAATTT 419
QY	838 TTCCGAGCGCTCGCTCAGGACAGCGCCCGCGGACAGCTTCAGTGGGCGCGCGCG 897
Db	
Db	418 TTTCGTAGTCGTGTTTAGGATACGGTTTTTCGGACGACGTTTTTAGTGGGGCTCGCG 359
QY	898 CCGTGC CGCGCGTCCCGCGGTTCCCGCGCGCTCCTCCCGCGCGCGCCTTCAGCGCGCTG 957
Db	
Db	358 TTTTGTTCGTCGTTGTTCGGTTTTTCGCGTTTTTTTTTCGCGCGTTTTTGTAGGGTTTTG 299
QY	958 CTCGCGCTCGCGGTACCGCGCGGCGAGCGCGCGGCTGCGCGCGCGCGCGCGCGCGAG 1017
Db	
Db	298 TTGTCGTTTTTCGCGTACGCGCGCGGCGATCGCGCGGTTTGGCGCGCGCGAGTGCAG 239
QY	1018 GTGCCACGACCGCGCGCCCTCTCTGCTGACCTCTCCGCGCGCGCGCCCGCCGCAAG 1077
Db	
Db	238 GTCGTATCGATCGCTCGTTTTTTTGTGTTATTTTTTCGCGCGCGGTTTTTCGTTAAG 179

Qy	1078	CCACTCGAGGCCGGCGGCGCGCGCACTGTACTGCCCCCTGCGGCTGCCGCA	1137
Db	178	TTATTTCGAGTTCGGCGGTTCGGCGCGGTATTGTATTGTTTTTTCGGTGTTCGTA	119
Qy	1138	GCCTCGAGCGGSCTTAGTCCGNCGTCTTGGCCCGCACTGTCGTACCCGTTGGAGACG	1197
Db	118	GTTTTGTAGCGGTTTTAGTTCGTCGTTTTGGTTCGTATTGTCGATTTCGGTGGAGACG	59
Qy	1198	CTCCTAGCTTGA	1209
Db	58	TTTTTTAGTTTGA	47

RESULT 6	
ABQ39024	
ID	ABQ39024 standard; DNA; 658 BP.
XX	
AC	ABQ39024;
XX	
DT	12-JUL-2002 (first entry)
XX	
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 25615..
XX	
KW	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW	drug; side effect; cancer; central nervous system; cardiovascular;
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;
KW	SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

01-SEP-2001:

01-SEP-2000: 2000DE-01043826-

XX  
PR 03-SEP-2000; Z000DE-U1044343:

PA (EPIG-) EPIGENOMICS AG.  
XX

PI Oleg A., Pierenbroock C., Verlin A., Guellu D,  
XX

DR WFI; 2002-3/1829/40.  
yy

DT Determining the degree of cytosine methylation

PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNPs); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ3410-CC ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention

W

Seq	Sequence	558 BP	59 A	108 C	222 G	269 T	0 U	0 Other
	Query Match	25.3%	Score	305.4	DB	6	Length	658
	Best Local Similarity	68.6%	Pred. No.	2.1e-32				
	Matches	420	Conservative	0	Mismatches	192	Indels	0
	Gaps	0						
Qy	598	GAGTACACCTTCGCCGACGAGGGGTCTTCGGCGCGCGCGAAGCGCTCAGCCACCGCGCG	657					
Db	1	GAGTATATTTTCGTGACGCGGGTTTTTCGTCGCTGCTAAGCGTTTAGTTATCGCGCG	60					
Qy	658	CCGGTCCCCCGCCCCCGGGTTCGGCCCGAGGAGCCCCCGGGCCCTCCGCCGCGCCCGCCG	717					
Db	61	TCGGTTTTTCGGTTCGGGTTCGGTTTCGAGAGAGTTTCGGGTTTTTTTCGTCGTTTCGTCG	120					
Qy	718	CCGCGCCCGCCGCCCGGCTCGCCCGCATCGCTCGCCCGCCCGCCAGGAGGACGC	777					
Db	121	TTCCGGTTCGTCGTTTCGGTTTCGTTTCGTAACGTTTCGTTTCGTTAGGAGGACGC	180					
Qy	778	GCCAGCCCCCGGGCAAGTTCCTCAGCTCTTCGCCATCGACATCTCTGCGCAAGCCC	837					
Db	181	GTTAGTTTCGCGGTAAATTTTTTTAGTTTTTCGTTATCGATAGTATTTTCGCTAAGTTT	240					
Qy	838	TTCCGACGCGTCGCTCAGGGACAGCGCCCCCGGGACGACGCTTCAGTCGGGCGCGCG	897					
Db	241	TTTCGTAAGTCGTGTTTTAGGATACGGTTTTTCGGACACAGTTTTAGTCGGGGCTCGCG	300					
Qy	898	CCCTGCCCGCGCTGCCCGGCTTCCCGCGCTCCTCCCGCGCGCCCTCGAGGGCCCTG	957					
Db	301	TTTTTGTTTCGCTGTTGTTTCGGGTTTTTCGCGTTTTTTTCGCGCGTTTTTCGAGGGTTTTG	360					
Qy	958	CTCGCGCTCTGCGGTACGCGCGGGGACGACCGCGCGGTTCGGCGCGCGGAGGCCGAG	1017					
Db	361	TTGTCGTTTTTCGCGTACGCGCGCGGGGAGTCGGCGGTTTGCGCGCGCGGAGTCGAG	420					
Qy	1018	GTGCAACCGACCGCGCCCTCTCTGTTGACCTCTCCCGCGCGCGGCCCGCCCAAG	1077					
Db	421	GTGTTATCGATCGCTGTTTTTTTTTGTTGTAATTTTTTCGCGCGCGTTTTTCGTTAAG	480					
Qy	1078	CCACTCGAGGCCGCGCGCGCGGGCGGACCTGTACTGTCCCCCTTCGCGCTGCCCGCA	1137					
Db	481	TTATTTTCAGGTTTCGGCGGTTCGGCGCGCGGATTTGTATTTGTTTTTCGCGTTTCGTA	540					
Qy	1138	GCCCTGACGCGCGCTTAGTCCGNCGTCCTGSGCCCGCACCTGTCGTACCCCGGTGAGACG	1197					
Db	541	GTTTTGTAGCGGTTTTAGTTCGTCGTTTTGTTGTTTCGTAATTTGTCGTAATTCGTCGAGACG	600					
Qy	1198	CTCCTAGCTTGA	1209					
Db	601	TTTTTAGTTTGA	612					

## RESULT 7

ADB75311

ID ADB

XX

AC ADB

XX

DT 04-11

XX

DE PRO

100  
 100  
 100  
 100

KW  
VY  
PRO:

XX OS HOM

US XX  
FROM

XX WO2  
PN

XX  
XX  
XX

PD 06-

XX

25-34

XX

PR 25-

PR 22-

PR 25-SEP-2001; 2001US-0325020P.  
 PR 12-DEC-2001; 2001US-0341746P.  
 PR 05-MAR-2002; 2002US-0362158P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;  
 PI Hoerh S, Kamatkar S, Wonsey AM, Glatt K, Zhao X, Anderson D;  
 XX WPI; 2003-248033/24.  
 DR New nucleic acid molecule, useful for diagnosing or treating prostate  
 PT cancer.  
 PT Disclosure; SEQ ID NO 135; 99pp; English.  
 PS  
 XX  
 XX The invention relates to newly discovered cancer markers associated with  
 CC the cancerous state of prostate cells. Also disclosed is a method of  
 CC assessing whether a patient is afflicted with prostate cancer. The method  
 CC of the invention involves assessing whether a patient is afflicted with  
 CC prostate cancer by comparing the level of expression of a marker in a  
 CC patient sample and the normal level of expression of the marker in a  
 CC control non-prostate cancer sample, where a significant increase in the  
 CC level of expression of the marker in the patient sample and the normal  
 CC level indicates that the patient is afflicted with prostate cancer.  
 CC Nucleic acids of the invention are useful for diagnosing or treating  
 CC prostate cancer, and may be useful in gene therapy. Sequences given in  
 CC ADB751177-ADB75631 represent marker cDNA and proteins. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 2487 BP; 413 A; 884 C; 782 G; 408 T; 0 U; 0 Other;  
 SQ  
 Query Match 16.7%; Score 201.4; DB 9; Length 2487;  
 Best Local Similarity 52.1%; Pred. No. 1e-18;  
 Matches 574; Conservative 0; Mismatches 512; Indels 16; Gaps 5;  
 QY 83 CCGCTCCCGCTGTGCGGGGGGAGAGAGACTCCCTGGGTCTAGATGGGACTGCGCGG 142  
 DB 517 CGACGACGAGGCGCGGTGGCGGGCCCGGGTGGCTCTCCCGGCGAGCGCGCGCGG 576  
 QY 143 CC--AAGCGCTCCGGGGGGGGCCAGAGATACGAGGCGGCGGCGGAGAGTGC 200  
 DB 577 GCGGCGCTGTACGCGGGGAGAGAGCTGAGGAGATCTGAGGAGGAGGAGGACGGA 636  
 QY 201 GGGAGCGGGCGCGCGGAGGAGCGGATCCGCGAGCAGCTGCTGAGCGGTGTGGC 260  
 DB 637 TGACATCTCTGTGCGCGCGCTGTGCGGGCTCCCGGCGCGCGCGCGCGCGCGG 696  
 QY 261 GGAGGCGGAGGCGGGCGCGGGCCAGGCGGCGGCGGCGGCGGCGGCGGCGG 320  
 DB 697 GCGGGGCGAGGCGCGGTGGGGCGG---CGGCGGCGGCGGCGGCGGCGGCGGAG 753  
 QY 321 TGCAGCGAGCAAGCCATATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 380  
 DB 754 CGCGGCTAGCGGCGCAAGAACCCGCTGTGAGCGCGGCTACTGCTATATCGGCTCAT 813  
 QY 381 CGCCTATGCGCATCCCGGAGCTGGCGGGGCGGCGCTTGACGCTGGCGGAGATCAAGAGTA 440  
 DB 814 CACTATGGCCATCTCTGCGAGAGCGCCCAAGAGCGGCTGACGCTGAGCGAGATCTGTGATT 873  
 QY 441 CCTCATGGCAAGTTCCCGCTTTTCCCGGCGGAGCTACAGCGGCTGGCGCAACTCCGTGCG 500  
 DB 874 CATCAGCGCGCTTCCCTACTACCGGAGAGTTCCTCCGCTGGCAGACACATCCG 933  
 QY 501 CCACAACTTTGCTCAACGACTGTTCTGTAAGTGTGCGGACCCCTCGCGCGCGCTG 560  
 DB 934 CCACAACTCTGCTCAACGACTGTTCTGTAAGATCCCGCGGAGCGCC---GGCAACCC 990  
 QY 561 GGGCAAGGACACTACTGATGCTCAACCCCAACAGCGAGTACACCTTCGCGCGAGCGGT 620  
 DB 991 GGGCAAGGGGCACTACTGAGCGCTGGACCGGAGTCCCGCGACATGTTTCGACACGCGAG 1050

QY 621 CTTCCGCGCGCGCGGCGGAGCGCTCAGCCACCGGCGGCGCGGCTCCCGCGCGCGGCTGCG 680  
 DB 1051 CTTCTCTGCGCGGAGGAGCGCTTCAAGCGGACGCGCTCTCCACCCCAACCGCGCGC 1110  
 QY 681 GCCCGAGGAGGCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 740  
 DB 1111 CGCGAGTCTCTGCTGCTGCGCGCGCGGAGCGGAGCGGCGCGCGCGCGCGCGCGCGC 1170  
 QY 741 GCCCGCATGCGTCCCGCTC 800  
 DB 1171 CGCGCGCGCGCTCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACG 1224  
 QY 801 CAGCTCTTCCGATCAGACAGCATCTTGGCGCAAGCCCTTCCGAGCGCGCTCGCTCAGGGA 860  
 DB 1225 CGGCGCTTACGCTGCGCTTACGCGCTTACGCGCTTACGCGCGCGCTTACGCGCGCT 1284  
 QY 861 CAGCGCGCGCGGAGCGGCTTCACTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 920  
 DB 1285 CTTCCGCGCGCGAGCTT 1344  
 QY 921 CCGCGCGCTCTCCCGCTT 980  
 DB 1345 GCCCGCTT 1404  
 QY 981 GGGCGAGCTT 1040  
 DB 1405 GCACCGCGCT--CGGCGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1462  
 QY 1041 CTTGCTTGACCTCTCCCG 1100  
 DB 1463 GGGCGCGCGCGCTCAGCGCTGCGCGCTGCGCGCTTCTCCATCGAGAGCATCATCGGG 1522  
 QY 1101 CGGCGCGCGCGCTTACTTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1160  
 DB 1523 AGCTTGGCT 1582  
 QY 1161 NCGTCTTGGCT 1182  
 DB 1583 TCGCGCTTGGCT 1604  
 RESULT 8  
 ID ABL68935 standard; DNA; 5181 BP.  
 XX  
 AC ABL68935;  
 XX  
 DT 15-MAY-2002 (first entry)  
 XX  
 DE Kidney cancer related gene sequence SEQ ID NO:7272.  
 XX  
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2001094629-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 30-MAY-2001; 2001WO-US010838.  
 XX  
 PR 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.

PR 22-SEP-2000; 2000US-0234567P.  
PR 25-SEP-2000; 2000US-0234923P.  
PR 25-SEP-2000; 2000US-0234924P.  
PR 25-SEP-2000; 2000US-0235077P.  
PR 25-SEP-2000; 2000US-0235082P.  
PR 25-SEP-2000; 2000US-0235134P.  
PR 25-SEP-2000; 2000US-0235280P.  
PR 26-SEP-2000; 2000US-0235637P.  
PR 26-SEP-2000; 2000US-0235638P.  
PR 27-SEP-2000; 2000US-0235711P.  
PR 27-SEP-2000; 2000US-0235720P.  
PR 27-SEP-2000; 2000US-0235840P.  
PR 27-SEP-2000; 2000US-0235863P.  
PR 28-SEP-2000; 2000US-0236028P.  
PR 28-SEP-2000; 2000US-0236032P.  
PR 28-SEP-2000; 2000US-0236033P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236109P.  
PR 28-SEP-2000; 2000US-0236111P.  
PR 28-SEP-2000; 2000US-0236422P.  
PR 29-SEP-2000; 2000US-0236891P.  
PR 02-OCT-2000; 2000US-0237172P.  
PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237294P.  
PR 02-OCT-2000; 2000US-0237395P.  
PR 02-OCT-2000; 2000US-0237316P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237606P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.  
PR 01-NOV-2000; 2000US-0245084P.  
XX  
PA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
PI WPI; 2002-198264/24.  
XX  
DR Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.  
XX  
PS Claim 1; SEQ ID NO 7272; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an anti-  
CC neoplastic agent. The method involves exposing cells to a chemical agent  
CC to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in AB461664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening an  
CC anti-neoplastic agent, and can be used for producing a product which is  
CC the data collected with respect to the anti-neoplastic agent as a result  
CC of M1, and the data is sufficient to convey the chemical structure and/or  
CC properties of the agent. M1 can be used in the treatment of cancer such  
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms  
CC tumour  
XX  
SQ Sequence 5181 BP; 1167 A; 1450 C; 1340 G; 1224 T; 0 U; 0 Other;  
Query Match 16.7%; Score 201.4; DB 6; Length 5181;  
Best Local Similarity 52.1%; Pred. No. 9.1e-19;  
Matches 574; Conservative 0; Mismatches 512; Indels 16; Gaps 5;  
XX 83 CGCCGTCCTGCTGGCGGCGGAGACGACTCCCTGGGCTCAGATGGGACTGCGCG 142

Db 3856 TCGCCCTCGCCTCGCCGGTGG 3877

RESULT 9

ADD29558

ID ADD29558 standard; mRNA; 1380 BP.

XX AC ADD29558;

XX DT 15-JAN-2004 (first entry)

XX DE Human tumour suppressor mRNA SEQ ID NO:7.

XX KW ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.

XX OS Homo sapiens.

XX PN WO2003058201-A2.

XX PD 17-JUL-2003.

XX PF 31-DEC-2002; 2002WO-US041825.

XX PR 31-DEC-2001; 2001US-0345317P.

XX PA (QUAR-) QUARK BIOTECH INC.

XX PA (CLEV-) CLEVELAND CLINIC FOUND.

XX PI Feinstein E, Gudkov AV;

XX DR WPI; 2003-598393/56.

XX PT Diagnosing cancer comprises determining the polypeptide or polynucleotide levels e.g., hepatic lipase, in a sample from a subject, where a higher level compared to that in a subject free of cancer is indicative of cancer.

XX PS Disclosure; SEQ ID NO 7; 272pp; English.

XX CC The invention relates to a novel method for diagnosing a cancer in a subject. The method comprises determining, in a sample from the subject, the level of at least one polypeptide, where a higher level of the polypeptide compared to the level of the polypeptide in a subject free of cancer is indicative of cancer. The polypeptide is selected from any of the polypeptides encoded by the polynucleotides listed in the specification and polypeptides which are at least 70% homologous to the polypeptides. The method of the invention has cytostatic activity, and may have a use in gene therapy. The method is useful in identifying markers specific for one or several types of cancer, depending on the tissue origin, which may be used in numerous diagnostic and prognostic applications as well as cancer type-specific targets for therapeutic intervention. The compounds that modulate the activity of a tumour suppressor gene are useful in the treatment of cancer or as anti-cancer drugs. The present sequence represents a polynucleotide of the invention.

XX SQ Sequence 1380 BP; 216 A; 496 C; 500 G; 168 T; 0 U; 0 Other;

Query Match 16.1%; Score 195; DB 9; Length 1380;

Best Local Similarity 53.1%; Pred. No. 8.3e-18;

Matches 533; Conservative 0; Mismatches 455; Indels 15; Gaps 5;

QY 2 TGAAGTTGAGGTGTCGTCCCTCGCGGGCCACCGGGACACAGAGGGCGATGACCTGG 61

Db 74 TGGACATCATGTGTGGCGGAGGCGACGCGGGCTGGAAGAGAGACAGCGACGCGAG 133

QY 62 AGGGGCGGGCGGACGACGCGCGCTCCCGG---CTGTGCGGCGCGGAGACGACTCCC 118

Db 134 GTTGCATAGCCCCCGGGGCGCGCGGAGCTGCGCTTGACGAGCGGACGAGGTGCCCC 193

QY 119 TGGGCTCAGATGGGAGCTCGCGGGCCAGCCCTTCGCGGGCGGGCGGCCCGCCAGATAGCG 178

Db 194 CGCGGCGCACCCCATCAGCGACAGCTCAGCCGCCGCCACCGACGAGCGCCCTGTGATGGCCCA 253

QY 179 AGGCGJACGGCGAAACAGAGTGCGGGAGCGCGCGCGCGGAGGAGGCGATCCCGGCGAG 238

Db 254 AGGAGGCGGCGGAGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 313

QY 239 CAGCTGCTGACGCGGTGTTGGCGGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGG 298

Db 314 GTGTCAAGGGCGGTGTTGGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 373

QY 299 GCGGCGGGGAGCGGCGAGGGTGCAGCGAGCAAGCCATATACGCGGCGGCGGCGGCGGCGG 358

Db 374 GCAGCGTTCGGCGGAGGCGCTGCCCCGAGCAAGCCCAAGAACAGAGCTAGTGAAGCGCG 433

QY 359 CCTACTCGTACATCGGCTCATCGCCATGCGCATCGCGACTCGGGGCGGCGGCGGCGGCG 418

Db 434 CTTACTCGTACATCGGCTCATCACCATGCGCATCTGTCAGAGCGGCGGCGGCGGCGGCGG 493

QY 419 CGTGGCGGAGATCAACGAGTACTCATGCGGCAAGTTTCCCTTTTCCCGCGGCGAGTACA 478

Db 494 CCCTGAGCGCATCTCGAGTTTCATCAGCAACCGCTTCCCTTACTACAGGAGAGTTC 553

QY 479 CGGCTGGCGCAACTCGTGGCGCACAACTTTGCTCAACGACTGCTTGTCAAGGTGC 538

Db 554 CCGCTTGGCAGAACAGCATCCGCCACAACTCTCACTCAACGACTGCTTGTCAAGATCC 613

QY 539 TGGCGGACCCCTCGCGCCCTGGGGCAAGGACAACACTACTGATGCTCAACCCCAACAGCG 598

Db 614 CCGCGGAGC---CGGCAACCCCGGCAAGGCACTACTGACCTTGGACCCCGCAGTCCG 670

QY 599 AGTACACCTTTCGCGGAGCGGGTTCCTCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGC 658

Db 671 AGGACATGTTTCGACAAACGCGAGCTTCTCGCGGCGCGGAAACGCTTCAAGCGCACCG 730

QY 659 CGGTCCCGCGCGCGGCTCGCGCCCGGAGGAGCGCGGCGCTCCCGCGCGCGCGCGCGCG 718

Db 731 AGAGCACTTTCGCGGAGCAGCGGCTCATGATGAGAGCTTCGGGCTTACAGCTGG 790

QY 719 CCGCGCGCGCGCGCGGCTTCGCCCGCATGCGCTTCGCCCGCGCGCGCGCGCGCGCGCG 778

Db 791 CGGCGGCGCGCGCGCGCGCGGAGCC---TACGCGCGCGCGCTTACGCGCTGCACCTGCG 846

QY 779 CAGACCCCGGCGGAGTTCCTCAGCTCTTCGCCATCGACAGC--ATCCTGCGCAAGCC 836

Db 847 GCGCGCGCGCGTGCCTATTTCGCAACCGGAGCGCGCGCGCGCGCGCGCGCTGCTGCGGCG 906

QY 837 CTTCCGCGAGCGCTCGGCTCAGGAGACAGCGCGCGCGCGCGCGCGCGCTTCACTGCGGCG 896

Db 907 CTCAGTACCGGTACGCGCTCGCGCGGTGGCACCGGTGCTGCTCCCGCTGTGCGCGC-- 964

QY 897 GCCTTGCCTCCCGCGCTGCGCGCGTTCGCCGCTCTTCGCCGCGCGCGCTTCAAGGCGCCT 956

Db 965 -TGCTGCGCTTCGGCGGAGTGGGCGCGCAAGCGCGCGCGCTTCGGCTCAGAGCTCGGCGCG 1023

QY 957 GCTGCGCGCTCTGCGCGTACGCGCGCGCGCGAGCGCGCGCGCGCTG 999

Db 1024 GGCTGCGAGTGCAGCTCAATAGCTTGGGCGCGCGCGCGCGCG 1066

RESULT 10

ABX76288

ID ABX76288 standard; DNA; 3482 BP.

XX AC ABX76288;

XX DT 02-APR-2003 (first entry)

XX DE Lung cancer-associated polynucleotide #152.

XX KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;

XX KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;

XX KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;

XX KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;

XX KW interstitial pulmonary fibrosis; asthma; bronchiectasis.

OS Unidentified.  
 XX WO200286443-A2.  
 XX 31-OCT-2002.  
 XX  
 XX 18-APR-2002; 2002WO-US012476.  
 XX 18-APR-2001; 2001US-0284770P.  
 XX 10-MAY-2001; 2001US-0290492P.  
 XX 09-NOV-2001; 2001US-0319245P.  
 XX 13-NOV-2001; 2001US-0350666P.  
 XX 29-NOV-2001; 2001US-0334370P.  
 XX 12-APR-2002; 2002US-0372246P.  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 XX Aziz N, MurRAY R;  
 XX WPI; 2003-093161/08.  
 XX P-PSDB; ABUS6559.  
 XX  
 XX Detecting a lung cancer-associated transcript in a cell from a patient  
 XX for treating lung cancer, by contacting a biological sample from the  
 XX patient with a polynucleotide that exhibits increased or decreased  
 XX expression in lung cancer.  
 XX  
 XX Claim 22; Page 303-304; 453pp; English.  
 XX  
 XX The invention relates to a method for detecting a lung cancer-associated  
 XX transcript in a cell from a patient, comprising contacting a biological  
 XX sample from the patient with a polynucleotide that selectively hybridizes  
 XX to a sequence that is at least 80 % identical to a gene that exhibits  
 XX increased or decreased expression in lung cancer samples. Lung cancer-  
 XX associated polynucleotides and polypeptides are used for identifying a  
 XX compound that modulates a lung cancer-associated polypeptide, for  
 XX inhibiting proliferation of a lung cancer-associated cell to treat lung  
 XX cancer in a patient and for treating a mammal having lung cancer by  
 XX administering a modulatory compound identified. The methods are useful  
 XX for treating lung cancer, such as small cell lung cancer, non-small cell  
 XX lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
 XX emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
 XX hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
 XX bronchiectasis. The genes, polynucleotides and polypeptides are useful  
 XX for diagnostic purposes and as targets for screening for therapeutic  
 XX compounds that modulate lung cancer, such as antibodies. Sequences  
 XX ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the  
 XX invention  
 XX  
 XX Sequence 3482 BP; 716 A; 1137 C; 981 G; 647 T; 0 U; 1 Other;  
 XX  
 XX Query Match 16.1%; Score 194.6; DB 7; Length 3482;  
 XX Best Local Similarity 57.9%; Pred. No. 7.8e-18;  
 XX Matches 385; Conservative 0; Mismatches 274; Indels 6; Gaps 2;  
 XX  
 XX 143 CCAAGCCGTCGCGGGCGGGGCCAGAGATACGACAGGCGGCGGCGAACAAGTGGCG 202  
 XX Db 611 CCGGACAGCGCGGGGATCCAGAGCCGGGGTGGCGGACGCCCGCCATGACTGCCG 670  
 XX QY 203 GAGGCGCGCGCGGCGGAGGAGCGATCCCGGCGAGCGATGCTGCGAGCGGTGGTGGCG 262  
 XX Db 671 AGACGGGCCCGCCCGCGGACCGG--GAGTGTGTTGTTACGTAAGAGAGAGCGCG 727  
 XX QY 263 AGGGCGGAGCGCGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGAGGCGGTG 322  
 XX Db 728 GCGAGACGCGACGAGGCGCGGGTCCAGGGGAGGCGACGCGGCGCGGGCGGGCGGCG 787  
 XX QY 323 CACGACAGAGCCATATACGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 382  
 XX Db 788 GGGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 847  
 XX QY 383 CCATGGCCATCCGGACTTCGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCG 442

848 CCATGGCCATCGCGCACGCGCCCGAGCGCCCTCAGCTGGGCGGCATCTACAAGTTCA 907  
 443 TCATGGGCAAGTTCCTCTTTTTCGCGGCGACGTACAGGGGTGGCGCAACTCCCGTGGCC 502  
 908 TCACGAGCGGTTCCCTTCTACCGGACACACCCCAAAAGTGGCAGACAGCATCCGCC 967  
 503 ACAACCTTTCGCTCAAGACTGCTTGTCAAGGTGCTGCGGACCCCTTCGGCGGCTGGG 562  
 968 ACAACCTCACACTCAAGACTGCTTCTCAAGATCCCGCGGAGCGCGGCGGCC--GG 1024  
 563 GCAGGACAACTACTGGATGCTCAACCCCAACAGGAGTACACCTTCGCGGACGGGTCT 622  
 1025 GTAAGGGCACTACTGGCGCTGACCCCAACAGGAGGACATGTTGAGAGCGGCGAGCT 1084  
 623 TCCGCGCGCGCAAGCGCTCAGCCACCGCGCGCGGTCCCGCGGCGCGGCGGCTGGCG 682  
 1085 TCCTGCGCGCGCAAGCGCTTCAAGCGCTCGGACCTTCACCTACCGCGCTTACATGC 1144  
 683 CCGAGGAGCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 742  
 1145 AGACGCGCGCGGCTCGCGCAGCGCGCGCTGCGGAGCGCGCGCGCGCGCGCGCGCG 1204  
 743 CCGCATCGCTCGCGCCCGCGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 802  
 1205 CCATCTTCCAGGCGCGTGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1264  
 803 GCTCC 807  
 1265 GTAC 1269

RESULT 11  
 ABZ35021  
 ID ABZ35021 standard; cDNA; 2271 BP.  
 XX  
 AC ABZ35021;  
 DT 05-FEB-2003 (first entry)  
 XX  
 DE Human gene expression profile polynucleotide SEQ ID NO 133.  
 XX  
 XX Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;  
 XX bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;  
 XX tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;  
 XX gene expression; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200274979-A2.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 20-MAR-2002; 2002WO-US008456.  
 XX  
 PR 20-MAR-2001; 2001US-0276947P.  
 XX  
 XX (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.  
 XX  
 PI Wan J, Wang Y;  
 XX  
 XX WPI; 2002-740862/80.  
 XX  
 XX New gene expression profile generated from primary, endothelial,  
 XX epithelial, and muscle cell types, useful for identifying disease  
 XX pathologies involving alterations of gene expression, e.g. cancer.  
 XX  
 PS Claim 3; Page 365-366; 850pp; English.  
 XX  
 XX The invention relates to a gene expression profile comprising one or more  
 XX genes (ABZ34899-ABZ35692) and generated from a cell type. The cell type  
 XX is a coronary artery endothelium, umbilical artery or vein endothelium,  
 XX aortic endothelium, dermal microvascular endothelium, pulmonary artery  
 XX endothelium, myometrium microvascular endothelium, keratinocyte

CC epithelium, bronchial epithelium, mammary epithelium, prostate  
CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,  
CC small airway epithelium, renal epithelium, umbilical artery smooth  
CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,  
CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,  
CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,  
CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast  
CC osteoblasts or prostate stromal cell. The gene expression profile is used  
CC for determining the level of RNA expression for a sample, determining the  
CC phenotype of a cell and distinguishing cell types. The gene or a protein  
CC expression profile is useful in identifying disease pathologies involving  
CC alterations of gene expression. The assessment of expression profiles may  
CC provide meaningful information with respect to tumour type and stage,  
CC treatment methods, and prognosis. The gene or protein expression profile  
CC may also be used for creating microarrays. The microarray is useful for  
CC genetic and physical mapping of genomes, DNA sequencing, genetic or  
CC medical diagnosis, genotyping of organisms, confirming cell or tissue  
CC identifications and in identifying promising antibiotics, antiviral or  
CC antifungal agents  
XX  
SQ Sequence 2271 BP; 392 A; 794 C; 705 G; 380 T; 0 U; 0 Other;  
  
Query Match 16.0%; Score 193.4; DB 6; Length 2271;  
Best Local Similarity 52.8%; Pred. No. 1.2e-17;  
Matches 490; Conservative 0; Mismatches 427; Indels 11; Gaps 3;  
  
QY 255 GGTGGGGAGGGCGCGAGCCGGGGCGCGGGCCAGCGCGGGCGGGGGAGCGG 314  
DB 440 GGCCCCGGGGGGGAGGACCGGTGGGGCGCGGGCGGGCGGGGGGG 499  
  
QY 315 CGAGGTGACGAGCAGCAGCATTATAGCGGGGGCCAGACCCCTACTCTACATGCG 374  
DB 500 CGGGAGCGGGGTAGCGGGGCCAAGAACCGGTGTGAAGCGGCCCTACTGTATATGCG 559  
  
QY 375 GCTCATCGCCATGCGCATCGGCACTCGGGGGCGGGCGGTGACGCTGGCGAGATCAA 434  
DB 560 GCTCATCACTATGGCCATCTGACAGAGCCCGAGAGCGGCTGACCTGAGCGAGATCTG 619  
  
QY 435 CGAGTACCTCATGGCAAGTTCCTTTTCGCGGAGCTACAGGGCTGGCGCAACTC 494  
DB 620 TGAGTTTATCAGCGGGCGGCTTCCCTACTACCGGGAGAAATTCCTCCGCTGGCAGAACAG 679  
  
QY 495 CGTGGCGCAACCTTTCGCTCAACGACTGCTTCGCAAGGTGCTGCGGAGCCCTCGCG 554  
DB 680 CATCCGCCCAACCTTCGCTCAACGACTGCTTCGCAAGATCCCGCGAGGCC---GG 736  
  
QY 555 GCCTCGGGCAGGACAACTACTGTGATGCTCAACCCCAACAGCGAGTACACCTTCGCCGA 614  
DB 737 CAACCCGGGCAAGGCAACTACTTGGAGCTTGGACCCGGAGTCCGCGACATGTTGACAA 796  
  
QY 615 CGGGGTCTTCGGCGCGCGCGAGCGCTCAGCACCGCGCGGGTCCCGCGCCCGG 674  
DB 797 CGGACGCTTCTCGCGCGGAGGAGCGCTTCAAGCGGAGCGCGCTCTCCCAACCAACGC 856  
  
QY 675 GCTCGGGCCGAGGAGCCCGGGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCC 734  
DB 857 CGCGCGCGCGAGTCTCTGCTGCTGCGCGCGCGGGAGCGCGAGGGCGCGCGGACCC 916  
  
QY 735 GGCTTCGCCCGCGATGGCTTCGCCCGCGCGCGAGAGAGCGCGCGCGCGCGCGGGA 794  
DB 917 GGCAGCGCGCGCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970  
  
QY 795 GTTCTCCAGTCTCTGCGCATCAGACGATCTTCGCGAAGCCCTTCGCGAGCGCTGCGCT 854  
DB 971 CGGCTAGGCGCCCTACGGCTGCGGCTACGGCTTGCAGCTTACGCGCGCGCCCTC 1030  
  
QY 855 CAGGAGACGCGCGCGCGAGAGCGCTTATGTTGGGCGCGCGCGCGCTGCGCGCGCTGCC 914  
DB 1031 GGCCCTCTTCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCCACCGCACTCGCC 1090  
  
QY 915 CGGTTTCCCGCGCTCTCTCCCGCGCGCGCTTCAGCGGCGCTTCTGCGCGCTGTCGCGTA 974  
DB 1091 CCGCTTTCGCTA 1150

QY 975 CGGCGGGGCGAGCGGGCGGGCTGGGGCGGGCGGGCGAGCGCGAGTGGCCACCGCGCGCC 1034  
DB 1151 CCGCGCCGACCCCGC--TCGGCGCGCGCCTTACC CGGCGCCCTTCCGCGGCCAAG 1208  
QY 1035 GCCCTCTCTGTTGCACTCTTCCCGGGGGGGCGCCCGCAAGCACTCCGAGGCCCGGC 1094  
DB 1209 GCGGGCGCGCGGGCGGCTCAGCGCTGCGGCTTCTCCATCGAGAGCATCATC 1268  
QY 1095 GCGCGGGCGCGGCACTGTACTGCCCTTGGGCTGCGGCTGCCGAGCCCTGAGGGGCGCTT 1154  
DB 1269 GGGGGCAGCTTGGGCGCGCGCGCTGCGCGCGCGCGCGAGCGCGCGCGCGCT 1328  
QY 1155 AGTCGCGNCGTCTCGCGCGCGCACTGTGCG 1182  
DB 1329 CAGGCTCGGCTCGGCTCGCGGTGCG 1356  
  
RESULT 12  
AAQ50632  
ID AAQ50632 standard; cDNA; 1860 BP.  
XX AAQ50632;  
XX  
DT 25-MAR-2003 (revised)  
DT 02-JUN-1994 (first entry)  
XX  
DE Brain factor-2.  
XX  
KW Brain factor; BF-1; BF-2; BF-3; DNA binding domain;  
KW brain transcription factor; diagnosis; tumour; cancer; probe;  
KW telencephalon; ss.  
XX  
OS Rattus rattus.  
XX  
PN W09323430-A1.  
XX  
PD 25-NOV-1993.  
XX  
PF 30-APR-1993; 93WO-US004102.  
XX  
PR 13-MAY-1992; 92US-00882292.  
XX  
PA (SLOK ) SLOAN KETTERING INST CANCER.  
XX  
PI Tao W, Lai E;  
XX  
DR WPI; 1993-386481/48.  
XX  
PT Isolated, animal nucleic acids encoding brain transcription factors -  
PT useful for diagnosis and treatment of abnormal brain factor synthesis in  
PT tumour tissue from animals and utilised as probe.  
XX  
PS Claim 2; Page 62-63; 96pp; English.  
XX  
CC The brain factors are transcription factors. Abnormal expression of BF-1  
CC in telencephalon-derived tissue or tumour tissue can be diagnosed. BF DNA  
CC or protein can be used to correct defective synthesis of BF. (Updated on  
CC 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 1860 BP; 402 A; 543 C; 545 G; 370 T; 0 U; 0 Other;  
  
Query Match 15.5%; Score 187.6; DB 2; Length 1860;  
Best Local Similarity 58.3%; Pred. No. 7.5e-17;  
Matches 394; Conservative 0; Mismatches 264; Indels 18; Gaps 3;  
  
QY 8 TGGAGGTGTTGTCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67  
DB 421 TCGACGTGTGGGGAGGGCGCGAGCGACGACGAGGAGGAGGAGCGATGACGACGAGGCG 480  
QY 68 CGGCGCGCGAGCGCGCGCTTCCCGCTGTCGCGCGGGGAGAGCACTCCCTGGGCTCAG 127  
DB 481 CGGCT 540





XX	Homo sapiens.	557	CCTGGGGCAAGGACAACTACTGATGTCTCAACCCCAACAGCGAGTACACCTTTCGCCGACG	616	
XX	OS				
XX	PN	WO9952415-A2.	911	C--GGCAAGGGCAGCTACTGGACGCTGGAGCCCGGACTCTTACAACTGTTTCGAGAAGC	967
XX	PD	21-OCT-1999.			
XX	PF	14-APR-1999; 99WO-US008159.	617	GGGTCTTCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGCGGTCCCGCGCGCGCGCGC	676
XX	XX		968	GCAGCTTCCTGGCGCGCGCGCGCTTCAAGAGAGAGACGCGGTGAAGGACAAGGAGG	1027
XX	XX	15-APR-1998; 98US-0081870P.	677	TGCGGCGCGAGAGAGGCC-----CGGGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGC	730
XX	XX	22-MAY-1998; 98US-00083351.	1028	AGAAGGACAGGCTGCACCTCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	1087
XX	PA	(IOWA ) UNIV IOWA RES FOUND.			
XX	XX	Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;	731	CCCCGCGCTCGCCCGCATGCGCTCGCGCGCGCGCGCGAGGAGCGCGCGCGCGCGCGC	790
XX	DR	WPI; 1999-620257/53.	1088	CGCGGAGCAGC	1147
XX	DR	P-PSDB; AAY43260.			
XX	PT	New isolated human forkhead transcription factor gene, FKHL7, used to	791	GCAAGTTCTCCAGCTCTTTCGCCCATCGACAGCATCTTCGCAAGCCCTTTCGCGAGCCGTC	850
XX	PT	develop products for the diagnosis, prognosis, monitoring, prevention or	1148	ACATCAAGACCGAGAACGGTACGTGCGCCCTCGCGCGCGCGCGCGCGCGCGCGCGC	1207
XX	PT	treatment of congenital heart disease.	851	GCCTCAGGGACACGGCCCGCGGACACGCTTCAGTGGGGCGCGCGCGCGCGCGCGCGC	910
XX	PS	Claim 11; Fig 1; 98pp; English.	1208	CCCTGGGCGAGCGGC	1267
XX	XX	This sequence encodes the human forkhead transcription factor gene,	911	TGCCCGCGTTCCCGCGCTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	970
XX	CC	designated FKHL7, of the invention. FKHL7 can be used in a novel method	1268	GCAGCTGTTCAGCGGGAGCAGCCCCCGGGCAGCTTCGCGTGGGGCGCGCGCGCTCAGCC	1327
XX	CC	for treating or preventing the development of a congenital heart disease			
XX	CC	(CHD) in a subject. The FKHL7 sequences can be used for diagnosis,	971	CGTACGGCGGGGAGCGCGCGCGCTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	1030
XX	CC	prognosis, monitoring, prevention and treatment of CHD. They can also be	1328	TGGACGCTGCGGATTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCACATA	1387
XX	CC	used for the production of transgenic animals and drug screening			
XX	XX	Sequence 3946 BP; 905 A; 1155 C; 1053 G; 833 T; 0 U; 0 Other;	1031	CGCCCGCCCTCTGCTTGCACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	1090
XX	XX	Query Match 14.5%; Score 175.6; DB 2; Length 3946;	1398	GCCAGGGCTTCAGCGTGGACAAACATCATGACGTGCTGCGGGGTTCGCGCAGAGCGCGG	1447
XX	XX	Best Local Similarity 50.1%; Pred. No. 2.5e-15;			
XX	XX	Matches 552; Conservative 0; Mismatches 534; Indels 16; Gaps 4;	1091	CGGC	1112
QY		23	CTCGC		
QY		372	CTCGC		
QY		432	GTGGGGGC		
QY		491	CCAAAGCGCTCGC		
QY		551	ACCAGC		
QY		611	TGTACTCGACCCCTGCGACCCCGAGCAGTACCCCGCGCGCGCGCGCGCGCGCGC		
QY		671	CCTACACGCGCAGC		
QY		731	TCATCAGCATGGCCATCCAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC		
QY		791	AGTACCTCATGGGCAAGTTTCCCTTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGC		
QY		851	TCCGCCAACACCTCTCGTCAACAGTGTCTGTCGTCGTCGTCGTCGTCGTCGTCGTC		

Search completed: March 30, 2004, 23:01:41  
Job time : 508 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 22:53:17 ; Search time 115 Seconds  
(without alignments)  
5834.223 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggagggtgttcgt.....tggagacgtcctagttga 1209

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	271.8	22.5	320	4	US-09-833-381-1266
2	187.6	15.5	1860	2	US-08-331-644-3
3	187.6	15.5	1860	5	PCT-US93-04102-3
4	184.4	15.3	1365	4	US-09-220-132-21
5	175.6	14.5	3946	3	US-09-083-351-1
6	175.6	14.5	3946	3	US-09-083-352-1
7	171.6	14.2	1659	3	US-09-083-351-3
8	171.6	14.2	1659	3	US-09-083-352-3
9	152.8	12.6	2830	1	US-07-882-292-1
10	152.8	12.6	2830	2	US-08-331-644-1
11	152.8	12.6	2830	5	PCT-US93-04102-1
12	134.6	11.1	1309	4	US-09-976-594-927
13	130.2	10.8	1155	2	US-08-331-644-4
14	130.2	10.8	1155	5	PCT-US93-04102-4
15	117.6	9.7	5080	4	US-09-976-594-435
16	110.6	9.1	1634	4	US-09-087-134-13
17	110.6	9.1	1793	3	US-09-113-309-1
18	110.6	9.1	1793	3	US-09-521-109-1
19	110.6	9.1	1793	4	US-09-563-332-1
20	104.4	8.6	2561	4	US-09-616-289-48
C 21	91.2	7.5	12001	1	US-08-458-568A-11
C 22	88	7.3	2561	4	US-09-616-289-48
C 23	86	7.1	4403765	3	US-09-103-840A-2
24	85.4	7.1	152331	3	US-09-128-155-16
25	85	7.0	12425	4	US-09-616-289-50
C 26	84.8	7.0	4411529	3	US-09-103-840A-1
C 27	83.6	6.9	4403765	3	US-09-103-840A-2

28	82.8	6.8	1605	4	US-09-087-134-10	Sequence 10, Appl
29	82.2	6.8	12001	1	US-08-458-568A-11	Sequence 11, Appl
30	81.6	6.7	697	3	US-09-040-384-17	Sequence 17, Appl
31	81.6	6.7	697	4	US-09-123-912-17	Sequence 17, Appl
32	81.6	6.7	697	4	US-09-643-597-17	Sequence 17, Appl
33	81.6	6.7	697	4	US-09-480-884A-17	Sequence 17, Appl
34	81.6	6.7	697	4	US-09-542-615A-17	Sequence 17, Appl
35	81.6	6.7	697	4	US-09-606-421B-17	Sequence 17, Appl
36	81.6	6.7	697	4	US-09-221-107-17	Sequence 17, Appl
37	81.2	6.7	4257	2	US-08-590-473-1	Sequence 1, Appl
38	81.2	6.7	4257	3	US-09-259-821A-1	Sequence 1, Appl
39	81.2	6.7	4257	3	US-08-843-659-1	Sequence 1, Appl
40	80.8	6.7	4411529	3	US-09-103-840A-1	Sequence 45, Appl
41	79.2	6.6	1614	4	US-09-616-289-45	Sequence 1, Appl
C 42	78.2	6.5	1578	1	US-08-681-129-1	Sequence 1, Appl
C 43	78	6.5	8438	1	US-07-945-283-1	Sequence 16, Appl
C 44	77.8	6.4	1668	4	US-09-087-134-16	Sequence 50, Appl
C 45	77.2	6.4	12425	4	US-09-616-289-50	

ALIGNMENTS

RESULT 1

US-09-833-381-1266  
; Sequence 1266, Application US/09833381  
; Patent No. 6672186  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs  
; FILE REFERENCE: 5800-119  
; CURRENT APPLICATION NUMBER: US/09/833,381  
; CURRENT FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 09/516,448  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 2050  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1266  
; LENGTH: 320  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(320)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-833-381-1266

Query Match		22.5%	Score 271.8;	DB 4;	Length 320;
Best Local Similarity		94.9%	Pred. No. 1.3e-35;		
Matches 296;		Conservative	0;	Mismatches 14;	Indels 2; Gaps 2;
Qy	567	GGACAATCTGCTGCTCAACCCCAACGAGTACACCTTCGCGACGGGGTCTTCG	626		
Db	1	GGACAATCTGCTGCTCAACCCCAACGAGTACACCTTCGCGACGGGGTCTTCG	60		
Qy	627	CCGCGCGCGCAAGCGCTCAACCCCAACGAGTACACCTTCGCGCGCGGGTGGGCGCGA	686		
Db	61	CCGCGCGCGCAAGCGCTCAACCCCAACGAGTACACCTTCGCGCGCGGGTGGGCGCGA	120		
Qy	687	GGAGGCGCGCGCTCCCG	745		
Db	121	GGAGGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	180		
Qy	746	GCATGCGCTCG	805		
Db	181	GCATGCGCTCG	240		
Qy	806	CTTTCG-CCATCAGCATCTCTGCGCAGCCCTTCGCGAGCGCTCGCTCAGGACAG	864		
Db	241	CTTTCNATCATCTACAGCATCTCTGCNCAATCCCTTCGCGATCGCGCGCTCANGACAG	300		
Qy	865	GCCCGCGGACGAC	878		





US-09-083-351-1  
; Sequence 1, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,351  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UIA-029.02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3946 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 475..2133  
; US-09-083-351-1

Query Match 14.5%; Score 175.6; DB 3; Length 3946;  
Best Local Similarity 50.1%; Pred. No. 2.9e-20;  
Matches 552; Conservative 0; Mismatches 534; Indels 16; Gaps 4;

Qy 23 CTCGCGCGCCACGGGGACAGCAGGCGAGTACTGAGAGGCGCGGCGCAGGACG 82  
Db 372 CTCGCGCGCGCGGACTCGGACTCGGCGCGCGCGCGCGCGCGCGCGCGG 431

Qy 83 CGCGCTCCCGCTGTGCGGCGCGGAGACACTCCCTGGGCTCAGATGGGACTGCGCG 142  
Db 432 GTGGGGGCG 490

Qy 143 CCAAGCGGTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 202  
Db 491 CCGTGTCCAGCCCACTCCCTCGGAGTGTGCGCTACCTCGCGCGCGCGCGCGCT 550

Qy 203 GAGCG 262  
Db 551 ACCG 610

Qy 263 AGGCGCGGAGCG 322  
Db 611 TGTACTCGACCTGCGACGCGGAGCAGTACCGCGCGCGCGCGCGCGCGCGCG 670

Qy 323 CACGACAGCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 376

Db 671 CCTACAGCGCGCAGCGCGCAGCCCAAGGACATGCTAGCGCGCGCTATAGCTACATCGCG 730  
Qy 377 TCATPGCCATGGCCATTCGCGCATCTGGCGCGCGCGCGCTTGAAGCTGGCGGATCAAG 436  
Db 731 TCATCACCATGGCCATTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 790  
Qy 437 AGTACCTCATGGGCAAGTTCCTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496  
Db 791 AGTTATCATGGACCGCTTCCCTTCTACCGGACACAGCAGGAGCTGCGAGACAGCA 850  
Qy 497 TGGCCCAACCTTTTGGCTCAAGCATGCTTCTCAAGGTGCTGGCGGACCGCTCGCGCG 556  
Db 851 TCGCCCAACCTTCTCGCTCAAGAGTGTCTTCTCAAGGTGCTGGCGGCGCGCGCGCG 910  
Qy 557 CTTGGGCAAGGACACTTACTGATGCTCAACCCCAAGCAGGAGTACACTTTCGCGCG 616  
Db 911 C---GGGCAAGGCGAGCTTCTGACGCTGACCGGACTCTTCAACATCTTTCGAGAG 967  
Qy 617 GGTCTTTCG 676  
Db 968 GCAGCTTCTCG 1027  
Qy 677 TGGCGCGCGGAGGCG 730  
Db 1028 AAGAGCAGGCTGACCTCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087  
Qy 731 CCGCGCGCTCG 790  
Db 1088 CGCGGAGGAGCG 1147  
Qy 791 GCAAGTCTTCAGCTCTCTGCGCATGACAGCATCTTGGCGCAAGCGCTTTCGCGAG 850  
Db 1148 ACATCAAGACCGGAGAGCGTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1207  
Qy 851 GCCTCAGGAGACACG 910  
Db 1208 CCTTGGGCG 1267  
Qy 911 TGCGCGGTTCCCGCGCGCTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970  
Db 1268 GCAGCTGTCTCAGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327  
Qy 971 CGTACG 1030  
Db 1328 TGGAGGTGCGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1387  
Qy 1031 CGCGCGCGCTCTGCTGTCACCTTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1090  
Db 1388 GCCAGGCGCTTACGCGTGGACAAATCATGCTGCTGCGCGCGCGCGCGCGCG 1447  
Qy 1091 CG 1112  
Db 1448 CCGCGAGCTCAGCTCGGCGCT 1469

RESULT 6  
US-09-083-352-1  
; Sequence 1, Application US/09083352  
; Patent No. 6207450  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS  
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston

```

STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,352
FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 475..2133
US-09-083-352-1

Query Match 14.5%; Score 175.6; DB 3; Length 3946;
Best Local Similarity 50.1%; Pred. No. 2.9e-20;
Matches 552; Conservative 0; Mismatches 534; Indels 16; Gaps 4

Qy 23 CTCGCGCGGCCACCGGGAGACAGCAGGSCATGACTCTGAGAGGGGCGGGCGGCGAGCG 82
Db 372 CTCGCGCGGGCCCGGACTCGGACTCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 431
Qy 83 CGCGCTGCCCGTGTGCGCGCGGGAGACGACTCCCTGGGCTCAGATGGGACTGCGCGG 142
Db 432 GTGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 490
Qy 143 CCAAGCGGTTCGCGGGCGGGCGGCGGACAGATACGACAGGCGGACGCGGACAGAGTGGG 202
Db 491 CGGTGTCCAGGCCCAACTCCCTGGAGTGTGCCCTACCTCGGCGGCGAGCAGAGCTACT 550
Qy 203 GAGCGGGGCGGGCGGGCGGAGGAGCGGATCCCGGCGAGCAGCTGTGCAGCGGTGGTGGCG 262
Db 551 ACGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 610
Qy 263 AGGGCGCGAGAGCGGGGCGGGCGGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGG 322
Db 611 TGTACTCGACCCCTGGCGACCGGACGATACCGGGGCGGATGGCCCGGCTACGGGC 670
Qy 323 CAGCGAGCAAGCCATATACGGGGCGGGCC-----AAGCCCCCTTACTGTATCTGCGC 376
Db 671 CCTACACGCGGACGCGGACGCCAAGGACATGTGTGAAGCGCGCCCTATAGCTACATCGC 730
Qy 377 TCATGCCATGCGCATTCGCGACTCGGCGGGCGGGCGCTTGAAGCTGGGAGATCAAGC 436
Db 731 TCATCACCATTGCGCATTCAGAACGCCCCGGAAGAGATCAACCTGGAACGCATCTACC 790
Qy 437 AGTACCTCATGGGCAAGTTCCTCTTTTCGCGGCGAGCTACAAGGCTGGCGCACTCGC 496
Db 791 AGTTTCATCATGACCGCTTCCCTCTTACCGGGACAAAGCAGGCGTGGCAGAACGACA 850
Qy 497 TCGCGCACACCTTTTCGCTCAACGACTGCTTCGTCAAGGTGCTGCGCGACCCCTCGCGC 556
Db 851 TCGCGCACAACTCTCGCTCAACGAGTGCCTTGTCAAGGTGCCCGCGGACGACAGAGC 910
Qy 557 CTTGGGCGAAGGACAACTACTGGATGCTCAACCCCAACAGCAGGATACACTTTCGCCGAC 616

```

[illegible]

```
QY 417 GACGTGGCGGAGATCAACAGTACTCTCATGGCAAGTTCCCTTTTCCGCGGAGTA 476
Db 297 CACCTCTGAACGCATCTACAGTTTCATGACCGCTTCCCTTCTACCGGCAACAA 356
QY 477 CACGGGCTGGCGAACTCGTGGCGGCAACCTTTTGGCTACAGACTGTTTGGTCAAGGT 536
Db 357 GAGGGCTGGCGAAGACAGATCCGCGCAACCTCTCGCTCAACGAGTGTCTTGGTCAAGGT 416
QY 537 GTGGCGGACCCCTCGCGCCCTGGGGCAAGACAACTACTGGATGCTCAACCCCAACAG 596
Db 417 GCGCGGAGCAACAAGACC---GGCAAGGCGAGCTACTGGAGCTGGACCGGAGTCT 473
QY 597 CGAGTACACTTTCGGGACGGGTCTTCCGCGCGCGCGCAAGCGCTCAGCCACCGGCG 656
Db 474 CTACAACATGTTTCGAGAACGGAGTCTTCTGCGCGCGCGCGGCGCTTCAAGAAAGGA 533
QY 657 GCGGTCTCCCGCGCGCGGCTCGCGCCCGGAGAGGCC-----CGGGCTTCCCGCGCG 710
Db 534 CGCGGTGAAGACAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 593
QY 711 CCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGAGTGGCTCGCGCGCGCGCG 770
Db 594 CGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 653
QY 771 GGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 830
Db 654 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 713
QY 831 CAAAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 890
Db 714 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
QY 891 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 950
Db 774 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833
QY 951 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1010
Db 834 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 893
QY 1011 GCGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1070
Db 894 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 953
QY 1071 CGCCAAGCGCATTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1112
Db 954 GGGGTGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 995
```

## RESULT 9

```
US-07-882-292-1
; Sequence 1, Application US/07882292
; Patent No. 5324638
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSES: John P. White
; STREET: c/o Cooper and Dunham, 30 Rockefeller
; STREET: Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/07/882,292
FILING DATE: 19920513
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41472
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 443..1882
OTHER INFORMATION:
FEATURE:
NAME/KEY: misc feature
LOCATION: 928..1255
OTHER INFORMATION: /note= "nucleotide sequence encoding DNA
OTHER INFORMATION: binding domain homology"
FEATURE:
NAME/KEY: misc signal
LOCATION: 1883..1885
OTHER INFORMATION: /note= "translation termination codon"
US-07-882-292-1
```

```
Query Match 12.6%; Score 152.8; DB 1; Length 2830;
Best Local Similarity 52.8%; Pred. No. 1.3e-16;
Matches 409; Conservative 0; Mismatches 347; Indels 18; Gaps 3;
```

```
QY 82 GCGCGCTCCCGCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141
Db 674 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 733
QY 142 GCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 201
Db 734 CGCGCTTCTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 793
QY 202 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 261
Db 794 AAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853
QY 262 GAGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 321
Db 854 AAGCGCGCGCGCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 913
QY 322 GCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
Db 914 GAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 973
QY 370 ATCGCGCTCATCGCCATGCGCCATGCGCCATGCGCCATGCGCCATGCGCCAT 429
Db 974 AAGCGCTCATCATGATGCGCCATGCGCCATGCGCCATGCGCCATGCGCCAT 1033
QY 430 ATCAAGAGTACCTCATGCGCGAGTTCCCTTTTCCGCGGAGAGTACAGCGGCTGGCG 489
Db 1034 ATCTAGAGTTTATCATGAGAACTTCCCTTACTACCGCGGAGAGAGAGAGAG 1093
QY 490 AACTCGGTGCGCGCAACCTTTTCTGCTCAAGAGTCTTCTGCTCAAGGTGCTGCGCGAGCC 549
Db 1094 AACTTCATCGCGCAACCTTCTGCTCAAGAGTCTTCTGCTCAAGGTGCTGCGCGAGTAC 1153
QY 550 TCAGCGCGCTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609
```

Db 1154 GACGACCC---GGGCAAGGCACTACTGATGCTGACCCGCTCGAGCGACGACGTGTC 1210  
Qy 610 GCGAGCGGGTCTTCGCGCCGCCCGGCAAGCGCTCAGCAGCGCGCGCTCCCGCG 669  
Db 1211 ATCGGCGGCAAGCGGCAAGCTGCGCGCGCTCCACCAAGCTTCGCGGCAAGTAGCC 1270  
Qy 670 CCGGGGTGCGGCGCGGAGGAGCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCG 729  
Db 1271 TTAAAGCGCGGCGACGGCTCACTCCACCGGCTCACCTTCATGACCGCG---CCGCG 1327  
Qy 730 GCGCGCGCTTCGCGCGCGATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789  
Db 1328 TCCCTCTACTGCGCGCGATGCGCGCTTCCTGCTCCCTGCAACCGCTCGCGCGCGCACT 1387  
Qy 790 GGCAAGTTCCTCAAGCTCTTCGCGCATCGACGAGCATCTGCGCAAGCGCTTCGCG 843  
Db 1388 TTGAGTTACAACGGGACCACTCGGCTACCCGAGCGACCCCATGCCCTACAGC 1441

## RESULT 10

US-08-331-644-1  
; Sequence 1, Application US/08331644  
; Patent No. 5976872  
; GENERAL INFORMATION:  
; APPLICANT: Tao, Wufan  
; APPLICANT: Lai, Eseng  
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/331,644  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/882,292  
; FILING DATE: 13-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 41472-A-PCT-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-278-0400  
; TELEFAX: 212-391-0525  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2830 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHEetical: N  
; ANTI-SENSE: N  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 443..1882  
; OTHER INFORMATION:  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 926..1255  
; OTHER INFORMATION: /note= "nucleotide sequence  
; OTHER INFORMATION: encoding DNA binding domain

; OTHER INFORMATION: homology"  
; FEATURE:  
; NAME/KEY: misc signal  
; LOCATION: 1883..1885  
; OTHER INFORMATION: /note= "translation termination  
; OTHER INFORMATION: codon"  
; US-08-331-644-1

Query Match Similarity 12.6%; Score 152.8; DB 2; Length 2830;

Best Local Similarity 52.8%; Pred. No. 1.3e-16;

Matches 409; Conservative 0; Mismatches 347; Indels 18; Gaps 3;

Qy 82 GCGCGCTCCCGCTGTCGCGCGCGGAGACGACTCCCTGGGCTCAGATGGGACTGGCG 141

Db 674 GCGCGCGAGCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 733

Qy 142 GCCAAGCGCTCCG 201

Db 734 CCGCTTCTGCTCCCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 793

Qy 202 GGAGCG 261

Db 794 AAGCGAGCG 853

Qy 262 GAGCG 321

Db 854 AAGCG 913

Qy 322 GCAGCGAGCAAGCCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369

Db 914 GAGCG 973

Qy 370 ATCGGCTCATCG 429

Db 974 AACCGCTCATCG 1033

Qy 430 ATCAACGAGTACCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 489

Db 1034 ATCTACGAGTTCATCATGAAGAACTTCCCTTACTACCGCGAGAACAGCGGCTGGCAG 1093

Qy 490 AACTCGGTGCG 549

Db 1094 AACTCCATCCG 1153

Qy 550 TCG 609

Db 1154 GACGACCC---GGGCAAGGCACTACTGATGCTCAACCCCAACAGCGAGTACACCTTC 1210

Qy 610 GCGAGCGGGTCTTCG 669

Db 1211 ATCG 1270

Qy 670 CCGGGCTGCG 729

Db 1271 TTAAAGCG---CCGCG 1327

Qy 730 GCGCGCGCTTCG 789

Db 1328 TCCCTCTACTGCG 1387

Qy 790 GGCAAGTTCCTCAGCTCTTCGCGCATCGACGAGCATCTGCGCAAGCGCTTCGCG 843

Db 1388 TTGAGTTACAACGGGACCACTCGGCTACCCGAGCGACCCCATGCCCTACAGC 1441

## RESULT 11

PCT-US93-04102-1  
; Sequence 1, Application PC/TUS9304102

; GENERAL INFORMATION:

; APPLICANT: Tao, Wufan

; APPLICANT: Lai, Eseng

; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF

```

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04102
FILING DATE: 19930430
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,292
FILING DATE: 13-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41472A-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 443..1882
OTHER INFORMATION:
FEATURE:
NAME/KEY: misc_feature
LOCATION: 936..1255
OTHER INFORMATION: /note= "nucleotide sequence encoding DNA
OTHER INFORMATION: binding domain homology"
FEATURE:
NAME/KEY: misc_signal
LOCATION: 1883..1895
OTHER INFORMATION: /note= "translation termination codon"
PCT-US93-04102-1

Query Match 12.6%; Score 152.8; DB 5; Length 2830;
Best Local Similarity 52.8%; Pred. No. 1.3e-16;
Matches 409; Conservative 0; Mismatches 347; Indels 18; Gaps 3;

QY 82 GCGCCCTCCCGTGTTCGGCGGGGAGAGAGACTCCCTGGGTCTAGATGGGACTGCCGG 141
DB 674 GCGCCGAGCCCCGCGAGCGCGCGCGCCCGAGCAGCGAGCGAGCAAGAGGCGCCCGCAG 733
QY 142 GCCAAGCGTCCGCGGGCGCGCGCCAGAGATACGCGAGGGGAGCGGCGAAGAGAGTGCG 201
DB 734 CGCTTCTGCTCCCGCGCTCCGCGCGCTTGAGCGGGCCCAAGGCTGACGCACTTGAGCC 793
QY 202 GAGGCGGGCGCGCGCGAGAGAGGCGATCCCGGACAGAGCTGTGTGACGCGTGTGTGGCG 261
DB 794 AAAGGCGAGCAGCGCGCGCGCTGCGGAGCTGGCGCGCGCGCGCGCGCGCGCGAGAGAG 853
QY 262 GAGGCGCGAGCGCGGGCGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 321
DB 854 AAGGCGCGGGCGCTGGGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 913
QY 322 GCACGCGCAAGCATTATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369

```

```

DB 914 GAGGGGGCAAGGAGGGCGCAAGAAACAACGCGCAAGTACGAGAGCGCGCTTCCACCTAC 973
QY 370 ATCGCGCTCATCGCATGGCCATCCGGGACTCCGGCGGGCGGGCGCTTGAAGCTGGCGGAG 429
DB 974 AACGGCTCATGATGGCCATCAGCAGAGTCCCGAGAAGCGCTGACGCTCAACGGC 1033
QY 430 ATCAACGAGTACCTCATGGGCAAGTTCCTTCCGCGGCGAGCTACAGGGGTGGCGC 489
DB 1034 ATCTACGAGTTTCATCAAGAACTTCCCTTACTACCGCGAGAAACAAGCAGGGGTGGCAG 1093
QY 490 AACTCCGTGGCGCCACAACCTTTCCTCAAGACTGCTTCTGTCAGAGTGTCTGCGCGACCCC 549
DB 1094 AACTCCATCCGCAACACTGTCCCTCAAGAGTCTTCTGTAAGGTACCGCGCCATAC 1153
QY 550 TCGCGGCGCTGGGGCAAGGCAACTACTGATGCTCAACCCCAACAGCGAGTACACCTTC 609
DB 1154 GACGACCC--GGGCAAGGGCACTACTGATGTGTGACCCGCTCGAGCGAGCGAGTTC 1210
QY 610 GCCGACGGGTCTTCGCGCGCGCGCAAGCGCTCAGCCACGCGCGCGGTCCCGCG 669
DB 1211 ATCGCGCGCAGCAGCGGCAAGCTCGCGGGCGCTCCACAGGTCTCGGCGCAAGCTAGCC 1270
QY 670 CCGCGGTGGCGCCGAGAGGGCGCGGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 729
DB 1271 TTAAAGCGCGGGCAGCGCTCACCTCCACGGGCTCACCTTCATGAGACCGCG--CGGCG 1327
QY 730 GCCCGGCGCTCGCGCGCGCATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789
DB 1328 TCCCTCTACTGGCGCGCATGTGCGCGCTTCTGTCTTCACACCCCTCGCGCGAGGAGCACT 1387
QY 790 GCGAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTCGCGCAAGCGCTTCCCGC 843
DB 1388 TTGAGTTACAAGGGGACCACTCGCGCTACCCGACGACCCCATGCGCTACAGC 1441

RESULT 12
US-09-976-594-927
; Sequence 927, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 927
; LENGTH: 1309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 3687719CBI
US-09-976-594-927

Query Match 11.1%; Score 134.6; DB 4; Length 1309;
Best Local Similarity 57.6%; Pred. No. 1e-13;
Matches 261; Conservative 0; Mismatches 189; Indels 3; Gaps 1;

QY 340 ACGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 399
DB 93 ACTGAGCAACCAAGCGCTCCCTACAGCTACATCGCGCTTATGCTATGGCGCATCCAGAGC 152
QY 400 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459
DB 153 TCACGGGGGCGAGCGGGCGCACCTCAGTGGGATCTACCGCTACATCATGCGCGGATTCGCC 212
QY 460 TTTTTCGCGCGAGCTACACGGGCTGGCGCAACTCCGTGCGCGCACAACTTTCGCTCAAC 519

```

Db 213 TTCTACCCACACACCGCGCGGCTGGCAGACAGATCCGCCACAACTGTCACTCAAC 272  
Qy 520 GACTGCTTCGTCAGTGTGCGGACCCCTCGCGGCGCTGGGGCAAGGACAACTACTGG 579  
Db 273 GAGTGTCTTGTCAAGGTGCGCGCGGATGACCGCAAGCC---AGGCAAGGCGAGTACTGG 329  
Qy 580 ATGCTCAACCCACACGAGTACACCTTCGCGGAGCGGGTCTTCGCGCGCGCGCGCAAG 639  
Db 330 AGCTGACCTGACTGCTCCAGCATGTTTGACACGCGAGCTTCTTACGCGCGCGCGCG 389  
Qy 640 CGCTCAGCCACCGCGCGCGGCTCCCGCGCGCGCGGCTGCGCGCGCGCGCGCGCGCG 699  
Db 390 CGCTTACCCCGGACAGAGTGTGAGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 449  
Qy 700 CTCGCCCG 759  
Db 450 CCCTCAGGCGGACCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 792  
Qy 760 GCGCGCGGAGGAGCG 792  
Db 510 TCATTCACCAGAGCTGCCAGATCCCAAGGCG 542

## RESULT 13

US-08-331-644-4  
; Sequence 4, Application US/08331644  
; Patent No. 5975872  
; GENERAL INFORMATION:  
; APPLICANT: Tao, Wufan  
; APPLICANT: Lai, Eseng  
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/331,644  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/882,292  
; FILING DATE: 13-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 41472A-PCT-US  
; TELEPHONE: 212-278-0400  
; INFORMATION FOR SEQ ID NO: 4:  
; LENGTH: 1155 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; US-08-331-644-4

Query Match 10.8%; Score 130.2; DB 2; Length 1155;  
Best Local Similarity 67.5%; Pred. No. 5.1e-13;  
Matches 199; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

Qy 352 AAGCCCCCTACTCGTATACGCGCTCATCGCATGSCCATCCGCGACTCGGCGGGGG 411  
Db 499 AAGCGCGCTACAGCTATATAGCGCTCATCACCATGGCGATCCAGAACGCGCCAGAGAAG 558  
Qy 412 CGCTGAGCGCTGGGGAGATCAACGAGTACCTCATGGGCAAGTTCCTTTTCCGGGCG 471  
Db 559 AAGATCACTCTGAACGGCATCTACCAGTTCATCATGAGCGGTTTCCCTCTTACCGGAG 618  
Qy 472 AGCTACAGCGGCTGGCGCACTCGTGGCGGACAACTTTTCGCTCAAGCACTGCTGTC 531  
Db 619 AACAGCAGGGCTGGCAGACAGCATCCGCCAACCTGTCTCAATGAGTGTGTTG 678  
Qy 532 AAGGTGCTGCGCGACCCCTCGCGGCGCTGGGGCAAGGACAACTACTGATGCTCAACCCC 591  
Db 679 AAGTGGCGCGGACGACAAAGAGCC---GGCAAGGCGAGTACTGAGCTCGACCG 735  
Qy 592 AACAGCGAGTACACTTCGCGGAGCGGGTCTTCGCGCGCGCGCGCGCGCGCTCA 646  
Db 736 GACTCTACAAACATGTTGAGATGGCAGCTTCTCGCGGCGCGCGCGCGCTTCA 790

## RESULT 14

PCT-US93-04102-4  
; Sequence 4, Application PC/TUS9304102  
; GENERAL INFORMATION:  
; APPLICANT: Tao, Wufan  
; APPLICANT: Lai, Eseng  
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John P. White  
; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/04102  
; FILING DATE: 19930430  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/882,292  
; FILING DATE: 13-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 41472A-PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-977-9550  
; TELEFAX: 212-664-0525  
; TELEX: 42523 COOP UT  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; PCT-US93-04102-4

Query Match 10.8%; Score 130.2; DB 5; Length 1155;  
Best Local Similarity 67.5%; Pred. No. 5.1e-13;  
Matches 199; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

Qy 352 AAGCCCCCTACTCGTATACGCGCTCATCGCATGSCCATCCGCGACTCGGCGGGGG 411  
Db 499 AAGCGCGCTACAGCTATATAGCGCTCATCACCATGGCGATCCAGAACGCGCCAGAGAAG 558

Search completed: March 31, 2004, 01:13:40  
Job time : 127 secs

QY 412 CGCTTGACGCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTTCGCGGC 471  
Db 559 AAGATCAGTCTGAACGGGATCTACAGTTATCATGGACCGTTTCCCTTCTACCGCGAG 618  
QY 472 AGCTACACGGGTGGCGCAACTCCGTGGCCCAACCTTTTCGCTCAACGAGTGTTCGTC 531  
Db 619 AACAGCAGGGGTGGCAGAACAGCATCCGCCCAACCTGTCACTCAATGAGTGTTCGTG 678  
QY 532 AAGGTGCTCGCGAGCCCTCGCGGCCCTGGGGCAGGACAACTACTGATGCTCAACCCC 591  
Db 679 AAGTGCCCGCGAGCAGCAAGAGCC--GGCAAGGCGAGCTACTGGACGCTGACCCG 735  
QY 592 AACAGCGAGTACACCTTTCGCGACGGGTCCTTCGCGCCGCCGCAAGCGCTCA 646  
Db 736 GACTCCTACACATGTTTCGAGAATGCGAGCTTCTTCGCGCGCGCGCGCTTCA 790

RESULT 15  
US-09-976-594-495  
; Sequence 495, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchdinger, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 495  
; LENGTH: 5080  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6673549 206866.1  
; NAME/KEY: unsure  
; LOCATION: 2468  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-976-594-495

Query Match 9.7%; Score 117.6; DB 4; Length 5080;  
Best Local Similarity 65.4%; Pred. No. 5.3e-11;  
Matches 204; Conservative 0; Mismatches 104; Indels 4; Gaps 2;  
QY 324 ACGCAGCAAGCCATATACGCGGCGGCCCAAGCCGCCCTACTCGTACATCGGCTCATCGC 383  
Db 22 ACGTCCAGCGCAGCTACCCGCGAGCGCCAGCGGCCCTACTCGTACATCTCGCTCATCAC 81  
QY 384 CATGGCCATCCGCGACTCGCGGGCGG-GCGCTTGACGCTGGCGGAGATCAACGAGTACC 442  
Db 82 CATGGCCATCCAGCAGCGGCCCGCCAGAGATGCTCACGCTGAGCGAGATCTACCAGTGA 141  
QY 443 TCATGGCAAGTTCCTCTTTTCGCGGCGAGTACACGGGTGGCGCAACTCCGTGGCGCC 502  
Db 142 TCATGACCTCTTCCCTATTACCGGCGAGACACAGCGCTGGCAGAACTCCATCCGCC 201  
QY 503 ACAACCTTTCGTCAACGACTCTCTTGTCAGGTGCTGCGGACCCCTCGCGGCCCTGGG 562  
Db 202 ACTCGTGTCCTCAATGACTGCTTCGTCAAGTGGCAGCGTCCCGCGCAAGCC--GG 258  
QY 563 GCAGGACACTACTGATGCTCAACCCCAACAGCGAGTACACCTTCGCGCGCGGGTCT 622  
Db 259 GCAAGGGCTCCTACTGAGCGCTGCACCCGGACTCCGCGCAACATGTTTCGAGAACGCGTCT 318  
QY 623 TCCGCCGCGGCC 634  
Db 319 ACTTGCGCGGCC 330

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      March 31, 2004, 00:21:47 ; Search time 446 Seconds
              (without alignments)
              10092.194 Million cell updates/sec

Title:       US-10-087-080-31
Perfect score: 1209
Sequence:    1 atgaagtggagggttcgt.....tggagagctcctagcttga 1209

Scoring table:  IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched:     2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database:      Published Applications NA.1

```

	Query Match	99.9%;	Score 1208;	DB 12;	Length 1209;
	Best Local Similarity	100.0%;	Pred. No. 8.9e-244;		
	Matches 1209;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY	1	ATGAAGTTGAGAGTGTTGTCCTCGCGCGGCCACAGGGGACAAAGCAGGCGCAGTGACCTG	60		
Db	1	ATGAAGTTGAGAGTGTTGTCCTCGCGCGGCCACAGGGGACAAAGCAGGCGCAGTGACCTG	60		
QY	61	GAGGGCGGGCGGCGAGCAGCGCGCGCTCGCGCGGGGAGACGACTCCCTG	120		
Db	61	GAGGGCGGGCGGCGAGCAGCGCGCGCTCGCGCGGGGAGACGACTCCCTG	120		
QY	121	GGCTCAGATGGGACTCGCGGGCCAAAGCGCTCGCGGGCGGGCGGCCACAGATACGCAG	180		
Db	121	GGCTCAGATGGGACTCGCGGGCCAAAGCGCTCGCGGGCGGGCGGCCACAGATACGCAG	180		

Result No.	Score	Query Match	Length	DB ID	Description
1	1308	99.9	1209	12	US-10-0229-345-9 Sequence 1
2	1308	99.9	1209	12	US-10-0274-177-9 Sequence 1
3	1308	99.9	1209	15	US-10-087-080-31 Sequence 1
C	412.6	34.1	585	15	US-10-027-632-231353 Sequence 1
5	271.8	22.5	320	9	US-09-833-381-1266 Sequence 1
6	201.4	16.7	2487	4	US-10-205-823-135 Sequence 1
7	193.4	15.0	2271	14	US-10-101-510-133 Sequence 1
8	184.4	15.3	2187	10	US-09-960-705-947 Sequence 1
9	161.6	13.4	1662	10	US-09-292-862-1 Sequence 1
10	161.4	13.3	2753	14	US-10-007-280A-115 Sequence 1
11	160.8	13.3	714	14	US-10-029-386-20688 Sequence 1
12	153.8	12.7	1506	9	US-09-963-285-9 Sequence 9
13	153.8	12.7	3289	9	US-09-963-285-8 Sequence 8
14	153.8	12.4	76458	9	US-09-963-285-1 Sequence 1
15	145.8	12.1	425	15	US-10-027-633-138703 Sequence 1



Db 661 GTCCCGCGCGCGGCTGGCGCCGAGAGAGCGCCCGGCTCCCGCGCGCCCGCGCGCC 720  
Qy 721 GCGCCCGCGCGCGCGGCTGGCGCCGAGAGAGCGCGCCGCGCGCGCGCGCGCGCGCGCC 780  
Db 721 GCGCCCGCGCGCGCGGCTGGCGCCGAGAGAGCGCGCCGCGCGCGCGCGCGCGCGCGCC 780  
Qy 781 AGCCCGCGCGCGGCAAGTCTTCAGCTCTTCGCGCATCGACATCTTCGCGCAAGCCCTTC 840  
Db 781 AGCCCGCGCGCGGCAAGTCTTCAGCTCTTCGCGCATCGACATCTTCGCGCAAGCCCTTC 840  
Qy 841 CGACGCGTGGCGCTCAGGAGACAGCGCCCGCGGAGAGCGCTTCAGTGGGGCGCGCGCGCC 900  
Db 841 CGACGCGTGGCGCTCAGGAGACAGCGCCCGCGGAGAGCGCTTCAGTGGGGCGCGCGCGCC 900  
Qy 901 TGCCCGCGCGCTGCGCGCTTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCC 960  
Db 901 TGCCCGCGCGCTGCGCGCTTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCC 960  
Qy 961 CGCTCTCGCGTACG 1020  
Db 961 CGCTCTCGCGTACG 1020  
Qy 1021 CCACGAGCG 1080  
Db 1021 CCACGAGCG 1080  
Qy 1081 CTCGAGCG 1140  
Db 1081 CTCGAGCG 1140  
Qy 1141 CTGAGCGCGCGCTTAGTCG 1200  
Db 1141 CTGAGCGCGCGCTTAGTCG 1200  
Qy 1201 CTAGCTTGA 1209  
Db 1201 CTAGCTTGA 1209

RESULT 3  
US-10-087-080-31  
; Sequence 31, Application US/10087080  
; Publication No. US20030235820A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Markowitz, Sanford David  
; APPLICANT: Eos Biotechnology, Inc.  
; APPLICANT: Case Western Reserve University  
; TITLE OF INVENTION: No. US20030235820A1el Methods of Diagnosis of Metastatic Colorectal  
; TITLE OF INVENTION: "Cancer, Compositions and Methods of Screening for  
; TITLE OF INVENTION: Modulators of Metastatic Colorectal Cancer  
; FILE REFERENCE: 018501-000840US  
; CURRENT APPLICATION NUMBER: US/10/087,080  
; CURRENT FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: US 60/272,206  
; PRIOR FILING DATE: 2001-02-27  
; PRIOR APPLICATION NUMBER: US 60/281,149  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: US 60/284,555  
; PRIOR FILING DATE: 2001-04-17  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 31  
; LENGTH: 1209  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: winged helix/forkhead transcription factor (HFH1)  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1161)  
; OTHER INFORMATION: n = g, a, c or t

US-10-087-080-31  
Query Match 99.9%; Score 1208; DB 15; Length 1209;  
Best Local Similarity 100.0%; Fred. No. 8.9e-244; Indels 0; Gaps 0;  
Matches 1209; Conservative 0; Mismatches 0

Qy 1 ATGAAGTTGAGAGTGTTCCTTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 1 ATGAAGTTGAGAGTGTTCCTTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Qy 61 GAGGGCG 120  
Db 61 GAGGGCG 120  
Qy 121 GGTCTCAGATGGGAGCTGCG 180  
Db 121 GGTCTCAGATGGGAGCTGCG 180  
Qy 181 GGCAGCG 240  
Db 181 GGCAGCG 240  
Qy 241 GCTGCTGCGAGCGTGTGCG 300  
Db 241 GCTGCTGCGAGCGTGTGCG 300  
Qy 301 GCG 360  
Db 301 GCG 360  
Qy 361 TACTCTGATGATGCGCTCATCG 420  
Db 361 TACTCTGATGATGCGCTCATCG 420  
Qy 421 CTGCGCGAGATCAACGAGTACCTCATGGGCAAGTTCCTTTTTCGCGCGCGCGCGCGCGCGCGCG 480  
Db 421 CTGCGCGAGATCAACGAGTACCTCATGGGCAAGTTCCTTTTTCGCGCGCGCGCGCGCGCGCGCG 480  
Qy 481 GCTGCGCGCAATCTCCG 540  
Db 481 GCTGCGCGCAATCTCCG 540  
Qy 541 CGCGACCCCTCG 600  
Db 541 CGCGACCCCTCG 600  
Qy 601 TACACCTTCG 660  
Db 601 TACACCTTCG 660  
Qy 661 GTCCCG 720  
Db 661 GTCCCG 720  
Qy 721 GCG 780  
Db 721 GCG 780  
Qy 781 AGCCCG 840  
Db 781 AGCCCG 840  
Qy 841 CGCAGCGCGTGGCGCTCAGGAGACAGCGCCCGCGGAGAGCGCTTCAGTGGGGCGCGCGCGCGCG 900  
Db 841 CGCAGCGCGTGGCGCTCAGGAGACAGCGCCCGCGGAGAGCGCTTCAGTGGGGCGCGCGCGCGCG 900  
Qy 901 TGCCCGCGCGCTGCGCGCTTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
Db 901 TGCCCGCGCGCTGCGCGCTTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
Qy 961 CGGCTCTGCGCGTACG 1020  
Db 961 CGGCTCTGCGCGTACG 1020



; APPLICANT: Schlegel, Robert  
 ; APPLICANT: Monahan, John E.  
 ; APPLICANT: Endege, Wilson O.  
 ; APPLICANT: Gannavarapu, Manjula  
 ; APPLICANT: Gorbacheva, Bella  
 ; APPLICANT: Hoersch, Sebastian  
 ; APPLICANT: Kamatkar, Shubhangi  
 ; APPLICANT: Monsey, Angela M.  
 ; APPLICANT: Glatt, Karen  
 ; APPLICANT: Zhao, Xumei  
 ; APPLICANT: Anderson, Dustin  
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
 ; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
 ; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER  
 ; FILE REFERENCE: MRI-044  
 ; CURRENT APPLICATION NUMBER: US/10/205,823  
 ; PRIOR FILING DATE: 2002-07-25  
 ; PRIOR APPLICATION NUMBER: 60/307,982  
 ; PRIOR FILING DATE: 2001-07-25  
 ; PRIOR APPLICATION NUMBER: 60/314,356  
 ; PRIOR FILING DATE: 2001-08-22  
 ; PRIOR APPLICATION NUMBER: 60/325,020  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: 60/341,746  
 ; PRIOR FILING DATE: 2001-12-12  
 ; PRIOR APPLICATION NUMBER: 60/362,158  
 ; PRIOR FILING DATE: 2002-03-05  
 ; NUMBER OF SEQ ID NOS: 455  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 135  
 ; LENGTH: 2487  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-205-823-135

Query Match 16.7%; Score 201.4; DB 14; Length 2487;  
 Best Local Similarity 52.1%; Pred. No. 2e-33;  
 Matches 574; Conservative 0; Mismatches 512; Indels 16; Gaps 5;  
 QY 83 CGCGTCCCGGTGCGCGCGGAGACGACTCCCTGGGCTCAGATGGGACTGCGCGG 142  
 DB 517 CGACGACGAGCGCGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 576  
 QY 143 CC--NAGCGTCCG 200  
 DB 577 GCGGCGTCTGACGCGCGCGGAGGACGAGCTGGAGGATCTGGAGGAGGAGGACG 636  
 QY 201 GGGAGCG 260  
 DB 637 TGACATCTGTGCG 696  
 QY 261 GGAGGCGCGGAGCG 320  
 DB 697 GCGGCGCGGAGCGCGGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753  
 QY 321 TGCAGCGAGAGCCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380  
 DB 754 CGCGGTHAGCG 813  
 QY 381 CGCGATGCGCATCCGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440  
 DB 814 CACTATGCGCATCTCTGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873  
 QY 441 COTCATGGGCAAGTTCCCTTTTCCGCGGAGCTACACGCGGCTGGCGCACTCCGTCG 500  
 DB 874 CATACGCGCGCTTCCCTTACTACCGGAGAGTTCCCGCGCGCGCGCGCGCGCGCG 933  
 QY 501 CCACAACTTTTGGTCAACGACTGTTCTGCTCAAGGTGTGCGCGACCCCTCGCGCGCG 560  
 DB 934 CCACAACTCTCTGCTCAACGACTGTTCTGCTCAAGATCCCGCGCGCGCGCGCG 990  
 QY 561 GGGGAGGACAACTACTGATGCTCAACCCCAAGCGAGTACACCTTTCGCGCGCGG 620

DB 991 GGGCAAGGGGAACACTACTGTGACCGTGGACCGGAGTCCGCGGACATGTTGACACAGCGAG 1050  
 QY 621 CTTTCCGCGCGCGCGCGCAAGCGCTCAGACACCGCGCGCGCGTCCCGCGCGCGCGCGTGG 680  
 DB 1051 CTTTCTGCGCGCGGAGGAAGCGTTCAAGCGGACGCGCTGCTCCCAACCGCGCGCG 1110  
 QY 681 GCGCGAGGAGGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 740  
 DB 1111 CGCGGAGTCTGTGCTGTGCGGCGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCG 1170  
 QY 741 GCGCGCATGCGCTCCCGCTTC 800  
 DB 1171 CGCGCGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 1224  
 QY 801 CAGTCTCTTGGCATCGACAGCATCTCGCAAGCCCTTCGCGAGCGCTGCGCTCAGGGA 860  
 DB 1225 CGCGCGCTACGCTGCGGCTGACGCTGCGCGCTTACGCGCGCGCTTACGCGCGCGCT 1284  
 QY 861 CACGCGCGCGCGGACGACGCTTCAAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCTG 920  
 DB 1285 CTTTCGCGCGCGGAGCT 1344  
 QY 921 CCGCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGG 980  
 DB 1345 GCTT 1404  
 QY 981 GGGCGAGCT 1040  
 DB 1405 GCACCGCT--CGGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1462  
 QY 1041 CTTGCTTGACCTTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1100  
 DB 1463 GCGCGCGCGCGCTCAGCGCTGCGCGCTGCGCGCTTTCATCGAGAGCATCATCGGG 1522  
 QY 1101 GCT 1160  
 DB 1523 AGCTTGGCG 1582  
 QY 1161 NGTCTGCG 1182  
 DB 1583 TCGCGCTCG 1604

## RESULT 7

US-10-101-510-133  
 ; Sequence 133, Application US/10101510  
 ; Publication No. US20030148295A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WAN, JACKSON  
 ; APPLICANT: WANG, YIXIN  
 ; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE  
 ; FILE REFERENCE: 15117.0012  
 ; CURRENT APPLICATION NUMBER: US/10/101,510  
 ; PRIOR FILING DATE: 2002-03-20  
 ; PRIOR APPLICATION NUMBER: 60/276,947  
 ; PRIOR FILING DATE: 2001-03-20  
 ; NUMBER OF SEQ ID NOS: 805  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 133  
 ; LENGTH: 2271  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-101-510-133

Query Match 16.0%; Score 193.4; DB 14; Length 2271;  
 Best Local Similarity 52.8%; Pred. No. 9.6e-32;  
 Matches 490; Conservative 0; Mismatches 427; Indels 11; Gaps 3;  
 QY 255 GGTGGCGGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 314  
 DB 440 GCGCGCGCGCGCGCGGCGGAGCGGTTGGGCGCGCGCGCGCGCGCGCGCGCG 499  
 QY 315 CGAGGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 374

500 CGGAGCGGGTACGGCGCCAAAGCCGCTGGTGAAGCCGCCCTACTCGTATATCGC 559  
375 GCTATCGCCATGCGCATCGCGACTCGCGGGCGGGCGCTTACGCTCGCGGAGATCAA 434  
560 GCTCATCATGATGCGCATCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619  
435 CGAGTACTCATGGGGAAGTTCCTCTTTTCGCGGCGAGCTACAGCGGCTGGCGAATC 494  
620 TGAGTTATCAGCGCGCGCTTCCCTACTACCGGAGAGATTCGCGCTGCGAGACAG 579  
495 CTGGGCGCAAACTTTTCGCTAAAGACTGCTTCTGCTAAAGGTCGCGGAGCCCTCGCG 554  
680 CATCGCGCAAACTTCTGCTCAAGACTGCTTCTGCTCAAGATCCCGCGAGCGCC---GG 736  
555 GCCTTGGGCGAAGAGAACTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 614  
737 CAACCCGGGCAAGGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 796  
615 CGGGGTCTTCCGCGCGCGCGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 674  
797 CGGAGGCTTCTGCGCGGAGGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 856  
675 GCTGGGCGCGAGAGCGCGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 734  
857 CGGCGCGCGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 916  
735 GGCTTGGCGCGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 794  
917 GCGAGCGCGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 970  
795 GTTCTCAGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 854  
971 CGGCTACGGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1030  
855 CAGGAGCAGCGCGCGCGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 914  
1031 GGCGCTTCTGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1090  
915 CGGTTCTCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 974  
1091 CCG 1150  
975 CGGCG 1034  
1151 CGGCG 1208  
1035 GCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1094  
1209 GCG 1268  
1095 GCG 1154  
1269 GCG 1328  
1155 AGTCCG 1182  
1329 CAGGCGCTGCG 1356

## RESULT 8

US-09-960-706-947  
; Sequence 947, Application US/09960706  
; Publication No. US20030134280A1  
; GENERAL INFORMATION:  
; APPLICANT: Mungen, William E.  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia  
; TITLE OF INVENTION: Gene Expression Profiles  
; FILE REFERENCE: 44921-5029-0105  
; CURRENT APPLICATION NUMBER: US/09/960,706  
; CURRENT FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: 60/223,323  
; PRIOR FILING DATE: 2000-08-07

; PRIOR APPLICATION NUMBER: 09/873,319  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 947  
; LENGTH: 2187  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 U13220  
US-09-960-706-947

Query Match 15.3%; Score 184.4; DB 10; Length 2187;  
Best Local Similarity 54.4%; Pred. No. 7.3e-30;  
Matches 492; Conservative 0; Mismatches 386; Indels 26; Gaps 5;

QY 278 GGGCGGCGGGCG 337  
DB 338 GCGCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 397  
QY 338 ATACGCGCGGGCG 397  
DB 398 TGGCGGCGGGCG 457  
QY 398 ACTCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 457  
DB 458 GCTCGGCG 517  
QY 458 CTTTCTTTCG 517  
DB 518 CTTTCTTTCG 577  
QY 518 ACGACTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 577  
DB 578 ACGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 634  
QY 578 GATGCTCAACCCCAACAGCGAGTACACCTTTCGCGCGCGCGCGCGCGCGCGCGCG 637  
DB 635 GGACCATGACCG 694  
QY 638 AGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 694  
DB 695 GCGGCTTTCAGCGGGAAGTGCAGCGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCG 754  
QY 695 GCGGCTTTCAGCGGGAAGTGCAGCGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCG 754  
DB 755 TGGGCTTTCG 814  
QY 755 GCG 814  
DB 815 CGCTCGGCTGCGCAAGCG 874  
QY 809 TCGGCATCGACAGCATCTTGGCAAGCCCTTCGCGAGCGCGCTTCGCGAGCACACGCG 868  
DB 875 ACGCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 934  
QY 869 CCGGAGCAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 922  
DB 935 GCGACATGTGCGCCCAACCGCGGTTCCACCTACATGCGCAGTGGCGCGCGCGCGCG 994  
QY 923 CCGCGCTTCTTCCCG 982  
DB 995 CCGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1054  
QY 983 GCGAGCG 1034  
DB 1055 GCAGCCCGGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1114  
QY 1035 GCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1094  
DB 1115 CGAGCCCTTGGCG 1174  
QY 1095 GCGCGGCG 1154

```
Db 1175 CCCTGACGCCAGCAGCAACCCCGCGCTCGGCGAGCGCTGCTCCAGCATGTCTCTCT 1234
QY 1155 AGTC 1158
Db 1235 ACTC 1238

RESULT 9
US-09-292-862-1
; Sequence, 1, Application US/09292862
; Publication No. US20030013087A1
; GENERAL INFORMATION:
; APPLICANT: Walter, Michael A.
; APPLICANT: Jordan, Tim
; APPLICANT: Raymond, Vincent
; TITLE OF INVENTION: NOVEL MUTATIONS IN THE FRA33 GENE FOR
; TITLE OF INVENTION: DIAGNOSIS AND PROGNOSIS OF GLAUCOMA AND ANTERIOR SEGMENT
; FILE REFERENCE: 07540/020003
; CURRENT APPLICATION NUMBER: US/09/292,862
; EARLIER FILING DATE: 1999-04-16
; EARLIER FILING DATE: 60/084,784
; EARLIER FILING DATE: 1998-05-08
; EARLIER FILING DATE: 60/082,206
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-292-862-1

Query Match 13.4%; Score 161.6; DB 10; Length 1662;
Best Local Similarity 51.3%; Pred. No. 4.5e-25;
Matches 483; Conservative 0; Mismatches 444; Indels 15; Gaps 4;

QY 183 CGACGCGCAACAGAGTGGCGGAGCGGCGCGCGCGCGAGGAGGCGATCCCGGAGCAGC 242
Db 57 CGCGCGGAGCAGAGTACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116
QY 243 TCGTGAGCGGTGGTGGCGAGGCGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 302
Db 117 CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 176
QY 303 CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
Db 177 CATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 236
QY 357 CCCCTACTGCTACATCGCGCTCATCGCCATCGCCATCGCGCGCGCGCGCGCGCGCGCG 416
Db 237 GCGCTATAGTACATCGCGCTCATCACCATGCGCCATCCAGAACGCGCGCGCGCGCG 296
QY 417 GACGCTGGCGGAGATCAACGAGTACTCATGGCAAGTTCCTCCCTTTTCGCGCGCGCG 476
Db 297 CACCTTGAAAGCGATCTACCAAGTTCATCATGAGCGCTTCCTCCCTTTTCGCGCGCG 356
QY 477 CACGCGCTGGCGCAACTCCGTCGCGCAACCTTTTCGCTCAACGAGTCTTCGTCAGAG 536
Db 357 GCAGGCGTGGCAGACAGCATCGCGCAACCTTCGCTCAACGAGTCTTCGTCAGAGT 416
QY 537 GCTGCGGACCCCTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 596
Db 417 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 473
QY 597 CGAGTACACTTTTCGCGGAGCGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
Db 474 CTACACATGTTTCGAGACAGCGAGTTCCTCGCGCGCGCGCGCGCGCGCGCGCGCG 533
QY 657 GCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
Db 534 CGCGTTGAGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 593
```

```
QY 712 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 770
Db 594 CGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 653
QY 771 GAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 830
Db 654 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 713
QY 831 CAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 890
Db 714 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
QY 891 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 950
Db 774 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833
QY 951 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1010
Db 834 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 893
QY 1011 GCGCGAGGTGCGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1070
Db 894 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 953
QY 1071 CGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1112
Db 954 GCGGTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 995
```

## RESULT 10

```
US-10-007-280A-115
; Sequence 115, Application US/10007280A
; Publication No. US20030059784A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; APPLICANT: Salceda, Susana
; APPLICANT: Chenghua, Liu
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0257
; CURRENT APPLICATION NUMBER: US/10/007,280A
; CURRENT FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,640
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 115
; LENGTH: 2753
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-007-280A-115
```

```
Query Match 13.3%; Score 161.4; DB 14; Length 2753;
Best Local Similarity 60.2%; Pred. No. 4.6e-25;
Matches 305; Conservative 0; Mismatches 196; Indels 6; Gaps 2;

QY 207 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 266
Db 256 CGAGCGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 315
QY 267 CGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
Db 316 GCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 375
QY 327 CAGCAAGCGCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 386
Db 376 GGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
QY 387 GCGCATCGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 446
Db 433 GCGCATCGCGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 492
```



QY 519 CGAGTCTTCTGCAAGTCTGCGGAGACCCCTCGCGGCCCTGGGCAAGGCAACTACTG 578  
Db 381 CGAGTCTTCTGCAAGTCTGCGGAGACCCCTCGCGGCCCTGGGCAAGGCAACTACTG 437  
QY 579 GATGCTCAACCCCAAGAGGAGTACACCTTCGCGGAGCGGGTCTTCGCGCGCGCGCA 638  
Db 438 GACCTTGAACCGGAGTCTTCAACATGTTGAGAACGCGAGCTTCTTCGCGCGCGCG 497  
QY 639 GCGCTCTACGCAACCGCGCGGCTCCCGCGCCCGGCTGCGGCCCGGAGGCGCCCGG 698  
Db 498 GCGCTTCAAAAGAGAGGAGTCTCAAGAGAGAGAGGAGGCGGCCACCTCAAGAGCC 557  
QY 699 CTTCCCG 758  
Db 558 GCG 617  
QY 759 CCGCGCGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 818  
Db 618 GCGCGAGAGAGAGGAGTCTCAAGAGAGAGAGGAGGCGGCCACCTAGCGGAGCG 677  
QY 819 CAGCATCTTGCACAAGCCCTTCGCGAGCGCTCGCTCAGGAGACAGCGCCCGCGGACG 878  
Db 678 CAAGTGGAGAGCGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737  
QY 879 GCTTCAAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 938  
Db 738 CAGCG 797  
QY 939 GCGCGCGTCAAGGCGCGCTCTCGCGCTTGCAGTACCGCGCGCGCGCGCGCG 995  
Db 798 GCTGCTGCTTCAAGTGGAGAAATATGATACCTTGGAGCTGCGCGCGCGGAG 857  
QY 996 GTGGGCGCGCGAGGCGGAGTGCACCGACCGCGCGCGCGCGCGCGCGCGCG 1055  
Db 858 GCTGAGCG 917  
QY 1056 CCG 1076  
Db 918 CCG 938

## RESULT 13

US-09-963-285-8  
; Sequence 8, Application US/09963285  
; Patent No. US2002090707A1  
; GENERAL INFORMATION:  
; APPLICANT: Enerbeck, Sven  
; APPLICANT: Krook, Katarina  
; APPLICANT: Rondahl, Lena  
; APPLICANT: Wasserman, Wyeth  
; TITLE OF INVENTION: PROMOTER SEQUENCES  
; FILE REFERENCE: 13425-042001  
; CURRENT APPLICATION NUMBER: US/09/963,285  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: SE 0004102-0  
; PRIOR FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: US 60/238,897  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: SE 0003435-5  
; PRIOR FILING DATE: 2000-09-26  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 3289  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-963-285-9

Query Match 12.7%; Score 153.8; DB 9; Length 3289;  
Best Local Similarity 52.4%; Pred. No. 1.7e-23;  
Matches 388; Conservative 0; Mismatches 347; Indels 6; Gaps 2;

QY 339 TAGCGCGCGCCCAAGCCCTTACTGTAATCGGGCTCATGCCATGGCCATCCGCGA 398  
Db 1397 TAGGACCTGGTGAAGCGCGCTTACAGTACATCGCGCTCATCACCATGCCATCCAGAA 1456  
QY 399 CTGCGCGCGCGCGCTTGAAGCTGGCGGAGATCAAGAGTACTCTATGGGCAAGTTCCC 458  
Db 1457 CGCGCGCGAGAGAAGATCACCTTGAAGGCACTTACAGTTTATATGACCGCTTCCC 1516  
QY 459 CTTTTTCCGCGCGAGCTTACACGGGCTGGGCAACTCGTGGCGCAACAACCTTTTCGCTCAA 518  
Db 1517 CTTTACCGGAGAGCAAGCAGAGGCTGGCAGAACATCCGCGCAACCTCTCGCTCAA 1576  
QY 519 CGACTGTTCTGTCAGAGTGTGCGGACCCCTCGCGGCCCTGGGGCGAGACAACTACTG 578  
Db 1577 CGAGTGTCTTCTCAAGGTGCGCGCGAGCAAGAAGCCCC--GGCAAGGCGAGTTACTG 1633  
QY 579 GATGCTCAACCCCAAGCAGGAGTACCTTCGCGCGAGCGGGTCTTCGCGCGCGCGCGCAA 638  
Db 1634 GACCTTGAGACCGGACTCTTCAACATGTTTCAGAGAGCGGAGTTCTCGCGCGCGCGG 1693  
QY 639 GCGCTCAGCAACCGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 698  
Db 1694 GCGCTTCAAAAGAGAGAGCGTGTCAAGGAGAGAGAGCGCGCGCGCGCGCGCG 1753  
QY 699 CTTCCCG 758  
Db 1754 GCG 1813  
QY 759 CGCGCGCGAGGAGCG 818  
Db 1814 GCGCGAGAGAGGTTGTTGATCAAGAGCGAGCGCGCTCCCGCGCGTCCCGGTCTAC 1873  
QY 819 CAGCATCTTCGCAAGCCCTTCGCGAGCGTCCCTCAGGAGACAGCGCGCGCGCGCG 878  
Db 1874 CAAGTGGAGACGCTGAGCGCGCGAGCGCGCTGCGAGGCGCGCGCGCGCGCGCG 1933  
QY 879 GCTTCAAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 938  
Db 1934 CAGCGCGCGCGCTCCCGCGAGCGTTCTGCTGCGGAGACCAACCGCGCGCGCGCG 1995  
QY 939 GCGCGCGCTGAGGCGCGCTGCTGCGGCTTCTGCGGTAAGCGCGCGCGCGCGCG 995  
Db 1994 GCTGCTGCTTCAAGCTGGAGAACATCATGACCTTCGGAACGTCGCGCGCGCGG 2053  
QY 996 GCTGGCGCGCGAGCGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1055  
Db 2054 GCTGAGCG 2113  
QY 1056 CCG 1076  
Db 2114 CCG 2134

## RESULT 14

US-09-963-285-1  
; Sequence 1, Application US/09963285  
; Patent No. US2002090707A1  
; GENERAL INFORMATION:  
; APPLICANT: Enerbeck, Sven  
; APPLICANT: Krook, Katarina  
; APPLICANT: Rondahl, Lena  
; APPLICANT: Wasserman, Wyeth  
; TITLE OF INVENTION: PROMOTER SEQUENCES  
; FILE REFERENCE: 13425-042001  
; CURRENT APPLICATION NUMBER: US/09/963,285  
; CURRENT FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: SE 0004102-0  
; PRIOR FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: US 60/238,897  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: SE 0003435-5  
; PRIOR FILING DATE: 2000-09-26  
; NUMBER OF SEQ ID NOS: 24

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6458
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2235)...(3737)
US-09-963-285-1

Query Match      12.78; Score 153.8; DB 9; Length 6458;
Best Local Similarity 52.44; Pred. No. 1.6e-23;
Matches 388; Conservative 0; Mismatches 347; Indels 6; Gaps 2;

Qy 339 TACGCGCGCGCCCAAGCCCTTACTCGTATACATCGCGCTCATCGCCATGCGCATCGCGCA 398
Db 2435 TAAGGACCTGTTAGAGCGCCCTTACAGTACATCGCGCTCATCAACATGCGCATCCAGAA 2494
Qy 399 CTGCGCGCGCGCGGCTTGAAGCTGGCGGAGATCAAGAGTACTCATGGGCAAGTTCC 458
Db 2495 CGCGCCCGAGAAGAAGATCACTTGAACGGCATCTACCAATTCATGACCGCTTCC 2554
Qy 459 CTTTTTCGCGCGGAGCTACACGGGCTGGCGCAACTCGTGGCGCACAACTTTCGCTCAA 518
Db 2555 CTTCACCGGAGACAAAGCAGGGCTGGCAGACAGCATCCGCCACAACTTCGCTCAA 2614
Qy 519 CGACTGCTTCGTCAGGTGCTGGCGACCCCTCGCGCCCTCGGGCCCTGGGCAAGCAACTACTG 578
Db 2615 CGAGTGTCTCGTCAAGTGTCCTCGCGACGACCAAGAGCC--GGCAAGGGCGAGTTACTG 2671
Qy 579 GATGCTCAACCCCAACAGCAGTACACTTTCGCGCAGGGGTCTTCGCGCGCGCGCA 638
Db 2672 GACCTGGACCGGACTCTTCAACATGTTGAGACAGGAGCTTCTCGCGCGCGGG 2731
Qy 639 GCGCTCAGCAACCGCGCGCGCTTCCTCGCGCGCGCGCTCGCGCGCGAGAGCGCCCGGG 698
Db 2732 GCGCTTCAAAAGAAAGACGTGTCAGGAGAGAGAGAGCGGGCCCACTCAAGGAGCC 2791
Qy 699 CTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCTCGCG 758
Db 2792 GCGCGCGCGCGCTCAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2851
Qy 759 CGCGCGCGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 818
Db 2852 GCGCGAGAAAGGTGTGATCAAGAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCTCATCAC 2911
Qy 819 CAGCATCTCGGCAAGCCCTTCGCGAGCGCTCGCTCAGGACACGCGCGCGCGCGCGCG 878
Db 2912 CAAGGTGAGACGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2971
Qy 879 GCTTACGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 938
Db 2972 CACGCGCGCGCGCTCCCGCGAGCGTTGCTGCGCGAGACACACGCGCGCGCGCGCGCG 3031
Qy 939 GCGCGCGCTGAGGCGCGCTGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 995
Db 3032 GCTGCTGCGTTCAGCGTGGAGAACATCATGACCTTCGGAACGTCGCCCGCGCGCGGAG 3091
Qy 996 GCTGGGCGCGCGGAGCGCGAGGTGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1055
Db 3092 GCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1076
Qy 1056 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1076
Db 3152 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3172
```

## RESULT 15

```
US-10-027-632-138703
; Sequence 138703, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138703
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138703
```

```
Query Match      12.11; Score 145.8; DB 15; Length 425;
Best Local Similarity 65.91; Pred. No. 1.1e-21;
Matches 226; Conservative 1; Mismatches 113; Indels 3; Gaps 1;

Qy 342 GCGCGCGCGCGCGCGCGCGCTACTCGTATCGCGCTCATCGCCATGGCCATCGCGACTC 401
Db 81 GCAGCGCGCAAGCGCCCTACTCTGATATCGCGCTCATCACCATGGCCATCTGCAAG 140
Qy 402 GCGCGCGCGCGCGCTTGAAGCTGGCGGAGATCAACAGTACCTCATGGGCAAGTTCCGCTT 461
Db 141 CCGGCACAAAGCGCTCAGCTCAGCGCATCTGCGCCTTCATTAGTGGCGCTTCCGCTA 200
Qy 462 TTTCCGCGGAGCTACAGCGGCTGGCGCACTCGGTGGCGCACACCTTTCGCTCAACGA 521
Db 201 CTACCGCGCGAAGTTCCCGCGCTGGCAGACAGCATCGCCCAACCTCTCGCTGAACGA 260
Qy 522 CTGCTTCTCAAGGTGCTGCGCGACCCCTCGCGCGCGCTGGGGCAAGGACAACTACTGAT 581
Db 261 CTGCTTCTCAAGATCCCGCGGAGC--CGGGCCACCCAGGCAAGGCAMCTACTGGAG 317
Qy 582 GCTCAACCCCAAGAGGATACACTTTCGCGGAGCGGGTCTTCGCGCGCGCGCGCGCGAGCG 641
Db 318 CTTGACCCCGCGCTCCCGAGGACATGTTTCGACATGGGAGCTTTCTCGCGCGGTAGGAAGCG 377
Qy 642 CTTGACCCCGCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 684
Db 378 TTTCAAGCGCCACCAACTGACCCCGGGAGCGCCACTGCGCCAC 420
```

Search completed: March 31, 2004, 02:26:07  
Job time : 450 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 22:45:27 ; Search time 2984 Seconds

(without alignments)  
12098.987 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggaggtgttcgt.....tggagacgtctcgttga 1209

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gas\_hum:\*

18: em\_gas\_inv:\*

19: em\_gas\_pln:\*

20: em\_gas\_vrt:\*

21: em\_gas\_fun:\*

22: em\_gas\_mam:\*

23: em\_gas\_mus:\*

24: em\_gas\_pro:\*

25: em\_gas\_rod:\*

26: em\_gas\_pig:\*

27: em\_gas\_vrl:\*

28: gb\_gsl:\*

29: gb\_gsl2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	537.6	44.5	904	13	BQ922461	BQ922461 AGENCOURT
2	350.8	29.0	559	9	AI169632	AI169632 EST215519
3	309	25.6	514	10	BF282916	BF282916 EST447507
4	301.2	24.9	398	9	AI586081	AI586081 vU92C05.x

```

c 5 269.2 22.3 425 10 BF920883
6 241.2 20.0 982 13 BQ30047
7 216.8 17.9 682 9 AL636071
c 8 215.2 17.8 518 10 BB637563
9 211.2 17.5 232 12 BI402320
10 209.8 17.4 564 13 BQ449254
11 209.8 17.4 946 14 CD754754
12 209.8 17.4 1117 14 CK028187
13 208.2 17.2 799 14 CF997203
14 205 17.0 805 13 EX084216
15 202.2 16.7 581 13 BQ285521
16 196.8 16.3 804 14 CK030218
c 17 183.2 15.2 688 12 BI443539
18 180.4 14.9 878 29 CN3044M2
19 179.2 14.8 246 29 CE688385
20 174.2 14.4 646 14 CB576131
21 170.8 14.1 542 14 CD282719
22 166.8 13.8 406 12 BM510453
23 163 13.5 885 11 BC019896
24 158 13.1 498 13 BX282141
25 158 13.1 635 14 CF125661
26 154.2 12.8 626 13 BY713753
27 153.4 12.7 842 12 BI254526
28 153.4 12.7 2293 11 BC055774
c 29 152.2 12.6 430 10 AW522942
c 30 152 12.6 445 10 AW236501
31 151.6 12.5 585 13 BQ480599
32 151.2 12.5 556 13 BQ189385
33 150.8 12.5 471 10 BE501220
c 34 150.4 12.4 2443 11 BC035020
35 149.6 12.4 356 12 BM529336
36 149.6 12.4 458 12 BM529896
37 149.6 12.4 471 12 BM529436
38 149.6 12.4 476 12 BM574346
39 149.6 12.4 483 12 BM529717
40 149.6 12.4 498 12 BM529729
41 149.6 12.4 541 12 BM572633
42 149.6 12.4 554 12 BM574305
43 149.6 12.4 556 12 BM573325
44 149.6 12.4 557 12 BM531626
45 149.6 12.4 557 12 BM532671

```

#### ALIGNMENTS

```

RESULT 1
BQ922461
LOCUS      BQ922461      904 bp      mRNA      linear      EST 20-AUG-2002
DEFINITION AGENCOURT 8963457 Lupski sciatic_nerve Homo sapiens cDNA clone
IMAGE:6200329.5', mRNA sequence.
ACCESSION  BQ922461
VERSION    BQ922461.1 GI:22337492
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 904)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs-rc@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA sequencing by: Agencourt Bioscience Corporation.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13615 row: a column: 02
            High quality sequence stop: 523.

```

```
FEATURES
Source      Location/Qualifiers
1..904
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6200329"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_hosts="DRI0B"
/clone_lib="Lupski_sciatic nerve"
/notes="vector: PCMV-SPORT6 (Life Technologies); Site 1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGCCG-3' and
5'-GACTAGTCTAGATCGGAGCGCCGCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
ORIGIN
Query Match      44.5%; Score 537.6; DB 13; Length 904;
Best Local Similarity 87.3%; Pred. No. 6.9e-52;
Matches 610; Conservative 0; Mismatches 87; Indels 2; Gaps 2;
QY 488 GCAACTCGTGGCGCAACACTTCGCTCAACGACTGCTTCGTCGAAGTGCTCGCGGACC 547
Db 1 GCAACTCGTGGCGCAACACTTCGCTCAACGACTGCTTCGTCGAAGTGCTCGCGGACC 60
QY 548 CCTCGCGGCTGGGCAAGGACAACTACTGATGCTCAACCCCAACAGCGAGTACACCT 607
Db 61 CCTCGCGGCTGGGCAAGGACAACTACTGATGCTCAACCCCAACAGCGAGTACACCT 120
QY 608 TCGCGGAGGGGCTCTTCGCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGCGTCCCG 667
Db 121 TCGCGGAGGGGCTCTTCGCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGCGTCCCG 180
QY 668 CGCCCGGCTGCGCGCGAGAGCGCCCGGCTTCGCCCGCGCGCGCGCGCGCGCG 727
Db 181 CGCCCGGCTGCGCGCGAGAGCGCCCGGCTTCGCCCGCGCGCGCGCGCGCGCG 240
QY 728 CGCGCGGCTGCGCGCGAGAGCGCTTCGCCCGCGCGCGCGCGCGCGCGCGCGCG 787
Db 241 CGCGCGGCTGCGCGCGAGAGCGCTTCGCCCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 788 CGGCAAGGTTCTCCAGTCTCTTCGCCATCGACAGCATCTTCGCAAGCCCTTCGCGAGCC 847
Db 301 CGGCAAGGTTCTCCAGTCTCTTCGCCATCGACAGCATCTTCGCAAGCCCTTCGCGAGCC 360
QY 848 GTGCTTCAGGACACGGCGCGCGAGAGCGCTTCAGTGGGCGCGCGCGCGCGCGCG 907
Db 361 CGCGCTTCAGGACACGGCGCGCGAGAGCGCTTCAGTGGGCGCGCGCGCGCGCGCG 420
QY 908 CGCTGCGCGCTTCGCCCGGCTCTTCGCCCGCGCGCGCTTCAGGCGCGCTTCGCGGCTCT 967
Db 421 CGCTGCGCGCTTCGCCCGGCTCTTCGCCCGGCTTCAGTGGGCGCGCGCGCGCGCG 480
QY 968 CGCGGTACGGCGCGGCGAGCGCGCGCTTGGCGCGCGCGAGCGCGAGTGGCACCGA 1027
Db 481 CGCGGTACGGCGCGGCGAGCGCGCGGCTTGGCGCGCGCGCGCGCGCGCGCGCG 540
QY 1028 CGCGCGCGCGCTCTGCTTCACCTTCGCCCGCGCGCGCGCGCGCGCGCGCGCG 1087
Db 541 TC-CCCGGCGCGCGGCTTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598
QY 1088 GCGCGCGCGCGCGCGCGCGAGCTGTACTGCCCGCTTCGCGCGCGCGCGCGCGCG 1147
Db 599 CCGATCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 658
QY 1148 CGCGCTTAGTCCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186
```

```
Db 659 TTGTTCCCGCCCTCCCGCTATCCCGCGCTTTGTTC 697

RESULT 2
A1169632/c
LOCUS
DEFINITION
EST215519 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
RKBV46 3' end, mRNA sequence.
ACCESSION
A1169632
VERSION
A1169632.1 GI:3709672
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Rattus sp.
REFERENCE
1 (bases 1 to 559)
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,W.D.
Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL
Unpublished (1998)
COMMENT
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
Location/Qualifiers
1..559
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2027517"
/db_xref="taxon:10118"
/clone="RKBV46"
/notes="Normalized rat kidney, Bento Soares"
/notes="Torgan: kidney; Vector: pT73Pac; Site 1: EcoRI;
Site 2: NotI"
ORIGIN
Query Match      29.0%; Score 350.8; DB 9; Length 559;
Best Local Similarity 80.2%; Pred. No. 1e-30;
Matches 451; Conservative 0; Mismatches 102; Indels 9; Gaps 3;
QY 590 CCAACAGAGGTATACACCTTCGCGAGCGGGTCTTCGCGCGCGCGCGCGCGCGCG 649
Db 559 CCAACAGAGGTATACACCTTCGCGAGCGGGTCTTCGCGCGCGCGCGCGCGCGCG 500
QY 650 ACCGCGCGCGGTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 709
Db 499 ACCGCGCGCGGTCTCCCGATCGCGGCTACGCGCGAGAGAGCGCGCGCGCGCG 440
QY 710 ----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 766
Db 439 GGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
QY 767 AGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826
Db 379 AGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 320
QY 827 TCGCAAGCGCTTCGCGAGCGGTTCGCGTTCAGGACACCGCGCGCGCGCGCGCGCG 886
Db 319 TCAGCAAGCGCTTCGCGAGCGGTTCGCGTTCAGGACACCGCGCGCGCGCGCGCG 260
QY 887 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946
Db 259 GGAGCGCTGTCTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 200
QY 947 GCAGGCGCGCTGTCTCTCTCGCGTTCGCGTTCAGGCGCGCGCGCGCGCGCGCG 1006
Db 199 GCGGTGCGCTGTCTCTCTGTGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 140
```



sequence: 5' CTCGAGT TTT TTT TTT TTT TTT TTT

```

sequence: 3 CAGCGAATTCCTTCTGCGCATCTCCTCGGCACAGCCCTTT 360

Query Watch      24.9%; Score 301.2; DB 9; Length 398;
Best Local Similarity 87.1%; Pred. No. 4e-25;
Matches 343; Conservative 0; Mismatches 48; Indels 3; Gaps 1;

QY    483 CTGGCGCAACTCCGTGCGGCCACAACCTTCGCTCAACGACTGTTCTGTCGAAGTGCTCGG 542
Db    1   CTGGCGCAACTCCGTGCGGCCACAACCTTCGCTCAACGACTGTTCTGTCGAAGTGCTCGG 60

QY    543 CGACCCCTCGCGGCCCTGGGGCAAGGACAIACTACTGATGCTCAACCCCAACAGCGAGTA 602
Db    61  CGACCCCTCGCGGCCCTGGGGCAAGGACAIACTACTGATGCTCAACCCCAACAGCGAATA 120

QY    603 CACCTTCGCGACGGGGTCTTCGCGCGCGCGCACAGCGCCTCAGCACCGCGCGCCCGGT 662
Db    121 CACCTTCGCGACGGGGTCTTCGCGCGCGCGCAAGCGCTCAGCACCGGACCACAGT 180

QY    663 CCCCGCGCCCGGGGTTCGGGCCCCGAGGAGGCCCGGGGCTCCCGCG---CGCCGCCGCC 719
Db    181 CTCGCGTCCGGGTTCGGGCGGAGGAAGCCCACCCGACCTGCCGGGACCCCGCAGCC 240

QY    720 CGCGCCCGCGCCCGGGCTCGCCCGCATCGCGTTCGCCGCCCGCGCAGGAGGACGGCG 779
Db    241 CGCGCCCGCGCCCGCTCTCTCCCGCATCGCGCGCTCGCGGGTCGCGCAGGAGGAGCGTC 300

QY    780 CAGCCCCCGGGCAAGTTCTCCAGTCTCTTCGCGCATCGACAGCATCTCTCGCGCAAGCCCTT 839
Db    301 CAGCCCTCGGAGCAAGTCTCTCAGTCTCTTCGCGCATCGACAGCATCTCTCAGCAAGCCCTT 360

QY    840 CGCGAGCGGTGCTCTCAGGACACGGCCCCCGGG 873
Db    361 TCGCAGCCCGCGCACCGCGACTCGGCTCTGGGG 394

```

```

RESULT 5
BF290883/c
LOCUS
DEFINITION
Rattus norvegicus cDNA clone RG1F65 3' sequence, mRNA sequence.
ACCESSION
BF290883
VERSION
BF290883.1 GI:11221953
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 425)
Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Pertea,G.,
Sultana.R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
Generation of ESTs from Normalized Rat Embryo, Bento Soares
Unpublished (2000)
Other ESTs: EST353025
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@igr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
FEATURES
Location/Qualifiers
1..425
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="RG1F65"
/tissue_type="mixed tissue"
/lab_host="DHS-alpha"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"

```

```

/note="Vector: pT3T7Pac; Site_1: EcoR1; Site_2: NotI;
Combination of ROV, RBR, RKI, RLI, RPL, RLU, REM, RMTU
RSP, RHE, RPC, RPN"

```

ORIGIN

Query Match 22.3%; Score 269.2; DB 10; Length 425;  
Best Local Similarity 80.1%; Pred. No. 1.7e-21;  
Matches 342; Conservative 0; Mismatches 79; Indels 6; Gaps 2;

QY 710 CCCCGCCGCGCGCGCGCGCGCGCTTCGCCGCATCGCTCGCCGGCCAGG 769  
DB 424 CCGCGAGCGCGCGCGCGCGCGCTTCCCACATCGCGGCTCGCCGCTCGCCAGG 365

QY 770 AGAGCGCGCGCACGCCCGCGGGGAAGTTCTCCAGCTCCTTCGCAATCGACAGCATCTCTGC 829  
DB 364 AGAGGGCTCCAGCGCGCGCGAGCAAGTTCTCCAGCTCCTTCGCAATCGACAGCATCTCTCA 305

QY 830 GCAGCGCTTCCGACGCGCTCGCTCAGGACACAGGCCCCGGGACAGCGTTCAGTGCGG 889  
DB 304 GCAAGCGGTTCGACAGCGCGCGGACGCGACCGGCTCTGGGGTGACACTACCCTGGA 245

QY 890 GCGCCGCGCCCTGCCCGCGCGCTCCCGCGCTTCGCCGCGCTCTCCCGCGCGCCCTTGCA 949  
DB 244 GCGCTGCTCCTGCCCGCGCGCTCGGCGCTATCCCGGCTCTTCCCGCGCTCGTCCGCGG 185

QY 950 GGCGCTCTGCGCGCTCTGCGCTAGCGCGCGCGGCGAGCGCGGCTGSGCGCGCGCGG 1009  
DB 184 GTGCGCTGCTGCGCTCTGTGCTTAGCGCGCGCGGAGCCACGCTGCTGGCTGCGCGG 125

QY 1010 AGCGCGAGGTGCGCACCGACCGCGCGCGCCCTCTCTGTGCACTCTCCCGCGCGCGCGCC 1069  
DB 124 GGGCGAGGTGCAAC--CGCGCGCGCCCTGTGCTGGCGCGCTCTCCACCGCGCGCC 68

QY 1070 CGCGCAAGCACTCCGAGCGCGG---CGCGCGCGCGCGCGCACTGCTACTGCCCCCTGCG 1156  
DB 67 CAGCCAAGCATTTCCAGGTCCGAGACCGCGCGCGCGCGCACTGCTACTGCCCCCTAC 8

QY 1127 GGCTGCC 1133  
DB 7 GGCTGCC 1

RESULT 6  
BQ430047  
LOCUS AGENCOURT\_7916258 NIH\_MGC\_68 Homo sapiens cdna clone IMAGE:6010412  
DEFINITION 5'', mRNA sequence.  
ACCESSION BQ430047  
VERSION BQ430047.1 GI:21169123  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 982)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-x@mail.nih.gov  
Tissue Procurement: DCTD/DTP/Gasdar  
cdna Library Preparation: Life Technologies, Inc.  
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LHAM13198 row: g column: 21  
High quality sequence stop: 491.  
Location/Qualifiers  
1..982  
/organism="Homo sapiens"  
/mol\_type="mRNA"

FEATURES  
source

```
/db_xref="taxon:9606"
/clone="IMAGE:6010412"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 68"
/notes="Organ: lung; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
```

## ORIGIN

```
Query Match      20.0%; Score 241.2; DB 13; Length 982;
Best Local Similarity 95.7%; Pred. No. 2.9e-18;
Matches 270; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

QY 1 ATGAAGTTGGAGTCTTCCTCGCGCGCCACGCGGACAGAGGAGTGAACCTG 60
DB 265 ATGAAGTTGGAGTCTTCCTCGCGCGCCACGCGGACAGAGGAGTGAACCTG 324

QY 61 GAGGCGCGCGGCGGAGCGCGCGTCCCGCTGTCGCGCGGCGGAGAGTCCCTG 120
DB 325 GAGGCGCGCGGCGGAGCGAGCGCGCGTCCCGCTGTCGCGCGGCGGAGAGTCCCTG 384

QY 121 GCTCAGATGGGAGTGGCGCGGCAAA---GCCGTCCGCGGCGCGGCGGAGAGTACG 177
DB 385 GCTCAGATGGGAGTGGCGCGGCAAAAGCGCGGCGCGGCGGCGGAGAGTCCG 444

QY 178 CAGGCGCGCGGCGGAGCGAGTGGCGGCGGCGGCGGCGGAGGAGGATCCCGCA 237
DB 445 CCGGCGCGGCGGAGCGAGTGGCGGCGGCGGCGGCGGCGGAGGAGGATCCCGCA 504

QY 238 GCAGCTG-CTGCAGCGGTGGTGGCGGAGGCGGCGGAGGCGCGG 278
DB 505 GCAGCTGCTGCAGCGGTGGTGGCGGAGGCGGCGGAGGCGGCGGTATGCGG 546
```

## RESULT 7

```
AL636071      682 bp mRNA linear EST 07-NOV-2003
LOCUS          AL636071
DEFINITION    AL636071 XGC-neurula Silurana tropicalis cDNA clone TNeu015112 5',
               mRNA sequence.
ACCESSION     AL636071
VERSION       AL636071.2 GI:38214106
KEYWORDS      Silurana tropicalis (western clawed frog)
SOURCE        Silurana tropicalis
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
               Xenopodinae; Silurana.
REFERENCE     1 (bases 1 to 682)
               Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
               Sanger Xenopus tropicalis EST project 2001 (11_2003)
               Unpublished (2003)
               On Nov 7, 2001 this sequence version replaced gi:16788050.
               Contact: Huckle E
               Sanger Institute
               Hinxton, Cambridgeshire, CB10 1SA, UK
               Email: trop@sanger.ac.uk
               This sequence is from a Xenopus Gene Collection (XGC) library
               constructed by Aaron M. Zorn.
               cDNA was oligo dt primed from Sug of poly A+ RNA from neurula.
               EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
               5' end and NotI at the 3' end.
               Vector: pCS107; Site_1: EcoRI; Site_2: NotI
               Host: Escherichia coli DH10B
               Sanger Xenopus tropicalis EST project 2001
               TROPICALIS SEQUENCE ID: TNeu015112.pikSP6
               Sequencing primer: SP6.
               Location/Qualifiers
```

## FEATURES

```
source
1..682
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
```

```
/clone="TNeu015112"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/clone_lib="XGC-neurula"
/notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."
```

## ORIGIN

```
Query Match      17.9%; Score 216.8; DB 9; Length 682;
Best Local Similarity 68.0%; Pred. No. 1.6e-15;
Matches 302; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 205 GGCGGCGCGGCGGAGGAGGCGATCCCGGAGAGCTGCTGCAGCGGTGGTGGCGGAG 264
DB 19 GGTAGCTTTGGGGGGATGAAGAAGATGAAGTGAAGAAGAGAGGAGGTAAACCCAGAG 78

QY 265 GGCGGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGGAGGTGCA 324
DB 79 AGAATGGGTCTTCGGCAGATGGTTCTACGCAATCCGAGCACAAATTGTTGAGGGGGG 138

QY 325 CGGAGCAAGCATATACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 384
DB 139 AAAACAAAGACTTACACCGCGCTTTTCGAAACCTCTTACTCTTACTCTTACTCTT 198

QY 385 ATGGCCATCGGACTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444
DB 199 ATGGCTATTAAAGACTCAGCCAGTGGCGGCGCTGACCCCTTGCAGAGATCAATGACT 258

QY 445 ATGGCAAGTTCCTCTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 504
DB 259 ATGAAGAAGTTCCTCTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 318

QY 505 AACCTTTTCGCTCAACGACTGCTTCGCAAGGTGCTCGCGGAGCCCTCGCGGCGGCGG 564
DB 319 AATCTTTTCCTCAACGACTGCTTCGCAAGGTGCTCGCGGAGCCCATCAGGCTTGG 378

QY 565 AAGCAACTACTGGATGCTCAACCCCAAGAGAGTACACCTTCCGCGGAGGGGTCTTC 624
DB 379 AAAGACAACACTACTGGATGCTGAACCCCAAGAGAGTACACCTTCCGAGAGCGGCTTC 438

QY 625 CGCGGCGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 648
DB 439 AGAGCGAGGAAGGAGCGACTTAAAC 462
```

## RESULT 8

```
BB637563/c     518 bp mRNA linear EST 26-OCT-2001
LOCUS          BB637563
DEFINITION     BB637563 RIKEN full-length enriched, adult male aorta and vein Mus
               musculus cDNA clone A53084C06 5', mRNA sequence.
ACCESSION      BB637563
VERSION        BB637563.1 GI:16473338
KEYWORDS       EST.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 518)
               Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
               Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
               Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
               Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
               Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
               Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
               Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.
               RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
               Unpublished (2001)
TITLE          Contact: Yoshihide Hayashizaki
               Laboratory for Genome Exploration Research Group, RIKEN Genomic
               Sciences Center(GSC), Yokohama Institute
               The Institute of Physical and Chemical Research (RIKEN)
JOURNAL
```



Research 6:791-806, 1996)  
TAG\_SEQ=None found"

## ORIGIN

Query Match 17.5%; Score 211.2; DB 12; Length 232;  
Best Local Similarity 94.4%; Pred. No. 6.1e-15;  
Matches 219; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 407 GCGGCGCTTACGCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTCC 466  
DB 1 GCGGCGCTGACGCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTCC 60  
QY 467 GCGGCGCTGACGCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTCC 526  
DB 61 GCGGCGCTGACGAGTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTCC 120  
QY 527 TCGTCAAGGTGCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTCC 586  
DB 121 TCGTCAAGGTGCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTCC 180  
QY 587 ACCCAACAGCGAGTACACCTTTCGCGGAGCGGGTCTTTCGCGGCGCGGCA 638  
DB 181 ACCCAACAGCGAGTACACCTTTCGCGGAGCGGGTCTTTCGCGGCGCGGCA 232

## RESULT 10

BQ449254 564 bp mRNA linear EST 29-MAY-2002  
LOCUS faa45a03.y1 zebrafish fin day3 regeneration Danio rerio cDNA clone  
IMAGE:5911949 5', similar to TR:070220 070220 FORK HEAD  
TRANSCRIPTION FACTOR. ;, mRNA sequence.

## ACCESSION

BQ449254

## VERSION

BQ449254.1

## SOURCE

EST.

## ORGANISM

Danio rerio (zebrafish)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 564)  
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,  
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterson, R., and Wilson, R.  
WashU Zebrafish EST Project 1998  
Unpublished (1998)

## TITLE

## JOURNAL

## COMMENT

Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu  
CDNA Library Preparation: Raymond Lee. CDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
ResourcenetZentrumPrimarDatenbank, Berlin, Germany (web address:  
www.rzpd.de)  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 450.  
Location/Qualifiers

## FEATURES

## SOURCE

1. 564  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:5911949"  
/sex="mixed male and female"  
/tissue type="3 day fin regenerates"  
/lab\_host="E. coli XL0LR"

/clone lib="zebrafish fin day3 regeneration"  
/notes"Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; 1st  
strand cDNA primed with (GA)10ACTAGTCTGAG(T)18, followed  
by second strand synthesis, and ligated to 5' adapter  
(5')-aattcgccagag-3', 3'-gccgtgctc-5'. cDNA was cloned  
directionally (EcoRI/XhoI) into Stratagene Zap express  
lambda phage arms. Mass invivo excision done to obtain  
inserts in pBK-CMV phagemid."

## ORIGIN

Query Match 17.4%; Score 209.8; DB 13; Length 564;  
Best Local Similarity 70.5%; Pred. No. 9.8e-15;  
Matches 280; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 251 CGGTGTGGCGGAGGCGCGGAGCGCGGCGCGGCGCGGCGCGGCGCGGCGGGA 310  
DB 94 CGGTGTGGCGGAGGAGGAGCTGGGTTCGGATGGAGACTCGGTGGCGCAGTCCGGCAC 153  
QY 311 GCGGCGAGGTGACGAGCAAGCCATATACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 370  
DB 154 CTGTGCGCGGACACCAAGGCAACCTTACACTCGGAGAGCCCAACCTCCATCTTACA 213  
QY 371 TCGCGCTCATCGCATGGCCATCGCGACTCGCGGCGGCGGCGGCGGCGGCGGCGGGA 430  
DB 214 TCGACTTATCGCATGGCCATCGGAGATCCCACTCCGCGCGGCGGCGGCGGCGGAAA 273  
QY 431 TCAACGAGTACCTCATGGGCAAGTTCCCTTTTCGCGGCGGCGGCGGCGGCGGCGGCA 490  
DB 274 TCAACGAGTACCTCATGAGAGAGTTCCCGTTTTTAGAGCGAGCTACACGGGTGGAGGA 333  
QY 491 ACTCGGTGGCGGCAACCTTTCGCTCAAGACTGCTTCGCTCAAGGTGCTCGGCGGCGGCG 550  
DB 334 ACTAGTGGCGGCAATCTGTCTTAAGAGCTGCTTCTCAAGGTCTTAAGGAGTCTCTT 393  
QY 551 CGCGGCGGCTGGGCGGAGGACAACTACTGGATGCTCAACCCCAACAGCGAGTACACCTTCG 610  
DB 394 CGAGACCGTGGGCGGAGGAGCAATTAAGTGTGAACCGCGGCGGCGGCGGCGGCGGCGG 453  
QY 611 CGAGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 647  
DB 454 CGGACGAGTGTTCGCGGAGGAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 490

## RESULT 11

## LOCUS

## DEFINITION

CD754754 946 bp mRNA linear EST 30-JUN-2003  
AGENCOURT\_14618637 NCI\_CGAP\_ZEMB2 Danio rerio cDNA clone  
IMAGE:6964741 5', mRNA sequence.

## ACCESSION

CD754754

## VERSION

CD754754.1

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

NIH-MGC http://mgc.nci.nih.gov/  
1 (bases 1 to 946)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Leonard I. Zon, M.D.  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM14605 Row: C Column: 12





Db 511 GTTGGAGAAATCGGTGCGCCACAACTGTGCACTTAACGACTGTTTTCTAAAGGTCTCTCC 570  
Qy 542 GCACACCCCTCGCGCCCTGGGCAAGACAACTACTGGATGCTCAACCCCAACAGAGACT 601  
Db 571 GGGACCCGTCAGACCCCTGGGAAAGACAACTACTGGATGCTGAATCTTCACAGCGACT 630  
Qy 602 ACACCTTCGCGCAGCGGGTCTTCGCGCGCGCGCGCAAGCGCTCA 646  
Db 631 ACACCTTCGCTGACGGAGTGTTCGACGACGAGGAAAGCGCATCA 675

RESULT 15

BQ285521 581 bp mRNA linear EST 14-MAY-2002  
faas1a09.y1 zebrafish fin day3 regeneration Danio rerio cDNA clone  
IMAGE:5912369 5' similar to SW:HFH1 RAT Q63244 HEPATOCYTE NUCLEAR  
FACTOR 3 FORKHEAD HOMOLOG 1 ; mRNA sequence.

ACCESSION BQ285521

VERSION BQ285521.1

KEYWORDS GI:20654021

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 581)

Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,

Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,

Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,

Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

Waterston,R. and Wilson,R.

WashU Zebrafish EST Project 1998

Unpublished (1998)

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

cDNA Library Preparation: Raymond Lee. cDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clone Distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T3 ET from Amersham

High quality sequence stop: 487.

Location/Qualifiers

1..581

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="IMAGE:5912369"

/sex="mixed male and female"

/tissue\_type="3 day fin regenerates"

/lab\_host="E. coli XL0LR"

/clone\_lib="zebrafish fin day3 regeneration"

/note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; 1st

strand cDNA primed with (GA)10ACTAGTCTCGAG(T)18, followed

by second strand synthesis, and ligated to 5' adapter

(5')-aattggcagcag-3', 3'-gccgtgttc-5'. cDNA was cloned

directionally (EcoRI/XhoI) into Stratagene Zap express

lambda phage arms. Mass in vivo excision done to obtain

inserts in pBK-CMV phagemid."

Qy 329 GCAAGCCATATACGCGGCGCCCAAGCCCCCTTACTCGTACATCGCGCTCATCGCCATGG 389  
Db 39 GCAAAACCCCTACACTCGGAGACCCAAACCTCCATACTTTATCATCGCACTTATCGCCATGG 98  
Qy 389 CCATCCGCGACTCGCGGCGGGCGGCTTGAGCTGGCGGAGATCAACGAGTACCTCATGG 448  
Db 99 CCATCCGAGACTCCAACTCCGCGCGAGTCACTTAGCGGAAATCAACGACTACTCATGA 158  
Qy 449 GCAAGTTCCTCCCTTTTTCGCGCGGAGCTACACGCGGCTGGCGCAACTCCGTGCGCCACAAC 508  
Db 159 AGAAGTTCCTCCCTTTTTAGAGGCGAGTACACCGGCTGGAGGAACTCAGTGGCCATAATC 218  
Qy 509 TTTGCTCAACGACTGCTTCTCAAGGTGCTGCGGACCCCTCGGCGCCCTGGGGCAAGG 568  
Db 219 TGTCTCTAAACGACTGCTTCTCAAGGTCTTAAAGGATCTTCGAGACCCGTGGGAAAGG 278  
Qy 569 ACAACTACTGGATGCTCAACCCCAACGCGAGTACACCTTCGCGGACGGGGTCTTTCGCCC 628  
Db 279 ACAATTACTGGATGCTGAACCCCGCACAGCGAGTACACCTTCGCGGACGGAGTGTTCGCA 338  
Qy 629 GCCGCGCAAGCGGCTCAG 647  
Db 339 GAAGGAGAAAGCGCATTAG 357

Search completed: March 31, 2004, 01:11:25  
Job time : 2989 secs

ORIGIN

Query Match 16.7%; Score 202.2; DB 13; Length 581;  
Best Local Similarity 77.1%; Pred. No. 7.2e-14;  
Matches 246; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 01:13:48 ; Search time 4679 Seconds  
(without alignments)  
11199.343 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggagtggttgtg.....tggagagcgtcctagcttga 1209

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pin.\*

35: em.htg.rod.\*

36: em.htg.man.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match %	Length	DB ID	Description
1	1209	100.0	1448	9	AF153341 Homo sapi
2	951	78.7	62761	9	AL499606 Human DNA
3	923	76.3	2322	9	BC053850 Homo sapi
4	923	76.3	4109	9	AF225950 Homo sapi
5	667	55.2	182152	2	AC067929 Homo sapi
6	554	45.8	178168	2	AC012350 Homo sapi
7	406	33.6	74969	2	AC016289 Homo sapi
8	146	12.1	182152	2	AC067929 Homo sapi
9	77	6.4	1513	10	AF154426 Mus muscu
10	77	6.4	2046	10	BC047155 Mus muscu
11	77	6.4	4763	10	AF010405 Mus muscu
12	77	6.4	187709	10	AL589738 Mouse DNA
13	53	4.4	1576	10	AF153193 Rattus ra
14	53	4.4	1760	10	RATHFH1
15	53	4.4	195184	2	AC119457
16	32	2.6	318	9	HSU42990
17	32	2.6	697	9	HSA337367
18	32	2.6	2011	9	AF275722 Homo sapi
19	32	2.6	106239	9	AL607122 Human DNA
20	32	2.6	153604	2	AC022754 Homo sapi
21	30	2.5	1038	5	AF072889 Xenopus l
22	30	2.5	1298	9	AF452723 Homo sapi
23	30	2.5	1817	5	XLAC011652
24	30	2.5	2068	9	AK122950 Homo sapi
25	30	2.5	2164	9	AY344639 Homo sapi
26	30	2.5	2207	9	AY345866 Pan trogl
27	30	2.5	2212	9	AY345859 Pongo pyg
28	30	2.5	2212	9	AY345860 Pongo pyg
29	30	2.5	2219	9	AY344641 Homo sapi
30	30	2.5	2219	9	AY344642 Homo sapi
31	30	2.5	2219	9	AY345862 Gorilla g
32	30	2.5	2219	9	AY345864 Pan trogl
33	30	2.5	2220	9	AY345861 Gorilla g
34	30	2.5	2487	9	AK125398 Homo sapi
35	30	2.5	3107	9	AF343005 Homo sapi
36	30	2.5	136659	9	AL772360 Human DNA
37	30	2.5	148489	2	AL773528 Homo sapi
38	30	2.5	158187	9	AL512605 Human DNA
39	30	2.5	164799	2	AC127362 Homo sapi
40	30	2.5	166636	2	AC140869 Homo sapi
41	30	2.5	172991	9	AL353608 Human DNA
42	30	2.5	174892	2	AL135795 Homo sapi
43	30	2.5	176734	9	HSA395114 Human DNA
44	30	2.5	176942	2	AC141462 Homo sapi
45	30	2.5	179342	9	BX255923 Human DNA

# ALIGNMENTS

RESULT 1  
AF153341  
LOCUS Homo sapiens winged helix/forkhead transcription factor (HFH1)  
DEFINITION Homo sapiens winged helix/forkhead transcription factor (HFH1)  
ACCESSION AF153341  
VERSION AF153341.1 GI:8489092  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1448)  
AUTHORS Hong,H.K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S.,  
Justice,M.J. and Chakravarti,A.

TITLE	The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in adult mouse
JOURNAL	Genes 29 (4), 163-171 (2001)
MEDLINE	21207067
PUBMED	11309849
REFERENCE	2 (bases 1 to 1448)
AUTHORS	Hong, H.-K., Noveroske, J.K., Justice, M.J. and Chakravarti, A.
TITLE	Direct Submission
JOURNAL	Submitted (21-WAY-1999) Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4955, USA
FEATURES	Location/Qualifiers
source	1..1448
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="6"
	/map="6p25"
gene	<233..>1441
	/gene="HFH1"
mRNA	<233..>1441
	/gene="HFH1"
	/product="winged helix/forkhead transcription factor"
CDS	233..1441
	/gene="HFH1"
	/codon_start=1
	/product="winged helix/forkhead transcription factor"
	/protein_id="AA575586.1"
	/db_xref="GI:8489093"
	/translations="MKLEVEVPRAAHKQGSDELAGGSDAPSPISAGDDSLGSDG DCAAKPSAGGARDTGDGQSGAGGFGAEAI,PA4AAAVVAEGAEAGAGGAGGAG GSGEARSKPTPTKPPYDYSYIALIAMAIRDAGRLTLAINEYLMGKFPFPRGSH GMNSVRLHSLNDPFCVKILRDSRPMGKNDYMWLNPSYTFADGVFRRRRKLSHR APVAPGLRPEEAPGLPAAPPAPAPASPMRSPASPAKVPAGKSSFAIDSL KRPFSRRRLDTAGTITLOWGAAACPLPAFPALLPAAPCRALLPLCAYGAGRPARGL ARAEVPPTPAPILLAPLPAAPAKPLRGPAAGGAHLYCPLRLPAAALQALVRRPGRH LSIPVETLLA"
ORIGIN	
Query Match	100.0%; Score 1209; DB 9; Length 1448;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1209; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATGAAGTTGAGGTGTTCTCCCTCGCGCGGCCACGGGACAGCAGGACAGGACGATGACCTG 60
DB	233 ATGAAGTTGAGGTGTTCTCCCTCGCGCGGCCACGGGACAGCAGGACGATGACCTG 292
QY	61 GAGGGCGCGCGCGCAGCAGCGGCCCTCCCGCTGTCGGCGCGGCGGAGACGACTCCCTG 120
DB	293 GAGGGCGCGCGCGCAGCAGCGGCCCTCCCGCTGTCGGCGCGGCGGAGACGACTCCCTG 352
QY	121 GGCTCAGATCGGAGCTCGCGGCCAACGCGTTCGGCGCGCGCGCGCGCAGATACGCAG 180
DB	353 GGCTCAGATCGGAGCTCGCGGCCAACGCGTTCGGCGCGCGCGCGCGCAGATACGCAG 412
QY	181 GGGCAGCGCGAAACAGAGTGGCGGAGGGCGGCGCGGCGCGGAGGAGCGCATCCCGGACGA 240
DB	413 GGCACGCGCGAAACAGAGTGGCGGAGGGCGGCGCGGCGCGGAGGAGCGCATCCCGGACGA 472
QY	241 GCTGCTCAGCGGTGTTGTCGGAGGGCGCGGAGCGCGGCGCGGCGCGGCGCGCGCGGCG 300
DB	473 GCTGCTCAGCGGTGTTGTCGGAGGGCGCGGAGCGCGGCGCGGCGCGGCGCGGCGCGGCG 532
QY	301 GGGCGGGGAGCGCGAGGGGTGCA CGCAGCAAGCATATACGGCGCGCGGCCCAAGCCCCC 360
DB	533 GGCAGCGGGAGCGCGAGGGGTGCA CGCAGCAAGCATATACGGCGCGCGGCCCAAGCCCCC 592
QY	361 TACTCGTACATCGGCTCATCGCCATGGCCATCGCGACTCGGGCGCGGCGCGCTTCACG 420
DB	593 TACTCGTACATCGGCTCATCGCCATGGCCATCGCGACTCGGGCGCGGCGCGCTTCACG 652
QY	421 CTGCGCGAGATCAACGAGTACTCTCATGGGCAAGTTCCTCTTTTCCTCGCGCAGCTACG 480
DB	653 CTGCGCGAGATCAACGAGTACTCTCATGGGCAAGTTCCTCTTTTCCTCGCGCAGCTACG 712

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/c-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> Rpl1-13J16 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone Rpl1-13J16 It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone Rpl1-13J16 is at 1 in this sequence. The true left end of clone Rpl4-668J24 is at 60762 in this sequence. The true right end of clone Rpl1-116B8 is at 4171 in this sequence.

## FEATURES

```

source
1..62761
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="6"
    /clone="Rpl1-13J16"
    /clone_lib="RPCI-11.1"
    /note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
17634..17641
    /note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
21165..21540
    /note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Assembly consistent with restriction digest data."
21466
    /note="tandem repeat. Forced join. Gap size estimated to be approximately 430bp by restriction digest data."
23824..24045
    /note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
60092
    /note="tandem repeat. Forced join. Gap size estimated to be approximately 120bp by restriction digest data."

```

## ORIGIN

```

Query Match      78.7%; Score 951; DB 9; Length 62761;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1001; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 152 CGCGCGCGCGCGCGAGATACGACGGGCGCGGCGGACACAGAGTGGGAGCGCGG 211
Db 17099 CGCGCGCGCGCGCGCGAGATACGACGGGCGCGGCGGACAGAGTGGGAGCGCGG 17158
QY 212 CGGCGCGCGGAGGCGGATCCCGGCGAGCAGTGTCTGACGGTGTGGCGGAGCGCGG 271

```

```

Db 17159 CGGCGCGGAGGAGGCGATCCCGGACAGTCTCTCTCAGCGGTGGTGGCGGAGCGCGG 17218
QY 272 AGCCCGCGCGCGCGCGGCGCGGCGCGGCGCGGCGGAGCGCGGAGGCGGAGCGAGCA 331
Db 17219 AGGCGCGCGCGCGCGGCGCGGCGCGGCGCGGCGGAGCGCGGAGGCGGAGCGAGCA 17278
QY 332 AGCATATACGGCGCGCGCGCGGCGCGGCGCGGCGCGGCGGAGCGCGGCGCGGCGCGCA 391
Db 17279 AGCATATACGGCGCGCGCGCGGCGCGGCGCGGCGCGGCGGAGCGCGGCGCGGCGCGCA 17338
QY 392 TCCGCGAGCTCGCGCGCGCGCGGCGCGGCGGCGGCGGAGATCAACGAGTACCTCATCGGCA 451
Db 17339 TCCGCGAGCTCGCGCGCGCGCGGCGCGGCGGCGGCGGAGATCAACGAGTACCTCATCGGCA 17398
QY 452 AGTTCGCGCGGCGCGGCGCGGCGGCGGCGGCGGAGTACCGGCGGCGGCGGCGGCGGCGG 511
Db 17399 AGTTCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGAGTACCGGCGGCGGCGGCGGCGGCGG 17458
QY 512 CGCTCAACGAGCTCTCTCGTCAAGGTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 571
Db 17459 CGCTCAACGAGCTCTCTCGTCAAGGTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 17518
QY 572 ACTACTGGATGCTCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631
Db 17519 ACTACTGGATGCTCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17578
QY 632 GCCGCAAGCGCGCTCAGCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691
Db 17579 GCCGCAAGCGCGCTCAGCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17638
QY 692 CCCCGGCGGCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 751
Db 17639 CCCCGGCGGCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17698
QY 752 GCTCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 811
Db 17699 GCTCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17758
QY 812 CCATCGACAGCATCTCTCGCGCAAGCGCTTCTCGGAGCGGCTCGCTCAGGCGACACGGCGCGCGG 871
Db 17759 CCATCGACAGCATCTCTCGCGCAAGCGCTTCTCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17818
QY 872 GGACGAGCTTCTAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 931
Db 17819 GGACGAGCTTCTAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17878
QY 932 TCCCGCGCGCGCGCGCTCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 991
Db 17879 TCCCGCGCGCGCGCGCTCGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17938
QY 992 CGCGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1051
Db 17939 CGCGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17998
QY 1052 CTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1111
Db 17999 CTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 18058
QY 1112 TGTAATGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1153
Db 18059 TGTAATGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 18100

```

```

RESULT 3
BC053850
LOCUS      BC053850
DEFINITION Homo sapiens forkhead box Q1, mRNA (cDNA clone MGC:61699
IMAGE:6010412), complete cds.
ACCESSION BC053850
VERSION   BC053850.1
KEYWORDS  MGC.
SOURCE    Homo sapiens (human)

```



Db 1168 CTGCCCGCGCTGCGCCGGTTCCTCCCGCGCTCTCTCCCGCGCGCCCCCTGCAGCGCGCTGTCT 1222

Qy 960 GCGCCTCTGCGCGCTACGCGCGCGGCGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGAGGT 1019

Db 1228 GCCGCTCTGCGCGTACGCGCGCGGCGAGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGAGGT 1287

Qy 1020 GCACCGACCGCGCGCGCTCTCTGCTTTGCACCTCTCCGCGCGCGCGCGCGCGCGCGCAAGCC 1079

Db 1288 GCCACCGACCGCGCGCGCTCTCTGCTTTGCACCTCTCCGCGCGCGCGCGCGCGCGCGCAAGCC 1347

Qy 1080 ACTCCGAGGCGCGCGCGCGCGCGCGCGCGCGCACCTGTACTTGCCCCCTGCGGCTGCCCGCAGC 1139

Db 1348 ACTCCGAGGCGCGCGCGCGCGCGCGCGCGCGCACCTGTACTTGCCCCCTGCGGCTGCCCGCAGC 1407

Qy 1140 CTTGCGAGCGCGCCT 1153  
|||||

Db 1408 CTTGCGAGCGCGCCT 1421  
|||||

RESULT 4

AF225950

LOCUS Hmo sapiens HNF-3/forkhead-like protein 1 (HFN1) gene, complete cds.

DEFINITION

ACCESSION AF225950

VERSION AF225950.1 GI:12655893

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 4109)  
Bieller,A., Pasche,B., Frank,S., Glaser,B., Kunz,J., Witt,K. and Zoll,B.  
Isolation and Characterization of the Human Forkhead Gene FOXQ1 DNA Cell Biol. 20 (9), 555-561 (2001)

TITLE

JOURNAL

PUBMED 11747606

REFERENCE 2 (bases 1 to 4109)  
Bieller,A., Pasche,B. and Zoll,B.  
Direct Submission

AUTHORS

TITLE Submitted (17-JAN-2000) Institute of Human Genetics, University of Goettingen, Heinrich-Dueker Weg 12, Goettingen 37073, Germany

JOURNAL Location/Qualifiers

FEATURES

source 1..4109  
          /organism="Homo sapiens"  
          /mol\_type="genomic DNA"  
          /db\_xref="taxon:9606"  
          /chromosome="6"  
          /map="6p23-25"  
          <1737..>2948  
          /gene="HFN1"  
          <1737..>2948  
          /gene="HFN1"  
          /product="HNF-3/forkhead-like protein 1"  
          1737..2948  
          /gene="HFN1"  
          /notes="forkhead transcription factor; similar to Mus musculus HNF-3/forkhead homolog 1 like encoded by GenBank Accession Number AF010405 and Rattus norvegicus HNF-3/forkhead homolog-1 encoded by GenBank Accession Number LI3201"

gene

mRNA

CDS

/codon\_start=1  
/product="HNF-3/forkhead-like protein 1"  
/protein\_id="AAK00639.1"  
/db\_xref="GI:12655894"  
          /translations="MKLEVFVPAAHGDKQSDLEGAGGDAPSLSAAGDDSLGSDGDCASNAPAGGARDPDGEGOSAGGRGAEEI PAAAAAVVAEGAAGAAGPAGNGASGEGARPKPYTRRPKPYSYIALAMIRADSAGRLTIATINMGKFPFFRGSYTGNRSVRNLNSLNDCFVKLRDPSPRKNYMLNPSEYTFADGVFRRKKLSHRAPVPAGLPPEEAGLPAPPPAAPAPAPRMRSPARQEERASPAGKFSSSFIDS LKPFPSRLRDTPQTLLQWGAACPPLPAPFALLPAPCPRALPLPCAYGAGERALGAREARVPPTAPILLILAPAAAAAPEKLPRGPAAGHAHYCLPLRLPALQAASVRRPFGHUPYPVEVTLLA"

QY 1140 CCTGCGAGCGGCT 1153  
 DB 2879 CCTGCGAGCGGCT 2892

RESULT 5  
 AC067929/c  
 LOCUS AC067929.2 182152 bp DNA linear HTG 24-AUG-2002  
 DEFINITION Homo sapiens chromosome 6 clone RP11-403120 map 6, WORKING DRAFT  
 SEQUENCE, 18 unordered pieces.  
 AC067929  
 ACCESSION AC067929.2 GI:8247824  
 VERSION HTG; HTGS PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS Birren, B., Linton, J., Nussbaum, C., Lander, B., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boukhalil, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melder, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission  
 JOURNAL Submitted (27-Apr-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 3 (bases 1 to 182152)  
 AUTHORS Birren, B., Linton, J., Nussbaum, C., Lander, B., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boukhalil, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melder, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission  
 JOURNAL Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Jun 4, 2000 this sequence version replaced gi:7655991.  
 All repeats were identified using RepeatMasker:

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information  
 Project name: L10103  
 Center clone name: 403.1.20  
 Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 17395 bases at least Q40  
 Consensus quality: 177417 bases at least Q30  
 Consensus quality: 179105 bases at least Q20  
 Insert size: 179000; agarose-fp  
 Insert size: 180452; sum-of-contigs  
 Quality coverage: 4.8 in Q20 bases; agarose-fp  
 Quality coverage: 4.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2222: contig of 2222 bp in length  
 2223: gap of 100 bp  
 2323: contig of 1463 bp in length  
 3785: gap of 100 bp  
 3786: gap of 100 bp  
 3885: gap of 100 bp  
 6753: contig of 2868 bp in length  
 6754: gap of 100 bp  
 6853: gap of 100 bp  
 10036: contig of 3183 bp in length  
 10136: gap of 100 bp  
 10137: gap of 100 bp  
 14537: contig of 4401 bp in length  
 14638: gap of 100 bp  
 21177: contig of 6540 bp in length  
 21178: gap of 100 bp  
 21277: gap of 100 bp  
 30240: contig of 8963 bp in length  
 30241: gap of 100 bp  
 40123: contig of 9783 bp in length  
 40223: gap of 100 bp  
 40224: gap of 100 bp  
 47999: contig of 7776 bp in length  
 48000: gap of 100 bp  
 58235: contig of 10136 bp in length  
 58335: gap of 100 bp  
 58336: gap of 100 bp  
 67042: contig of 8706 bp in length  
 67141: gap of 100 bp  
 75921: contig of 8780 bp in length  
 75922: gap of 100 bp  
 76022: contig of 10506 bp in length  
 86527: gap of 100 bp  
 86528: gap of 100 bp  
 101334: contig of 14707 bp in length  
 101335: gap of 100 bp  
 101435: contig of 16196 bp in length  
 117730: gap of 100 bp  
 117731: contig of 20563 bp in length  
 138293: gap of 100 bp  
 138393: gap of 100 bp  
 138394: contig of 19448 bp in length  
 157842: gap of 100 bp  
 157942: contig of 24211 bp in length.

Location/Qualifiers  
 1. 182152  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6"  
 /clone="RP11-403120"

FEATURES  
 source

```
/clone lib="RPCI-11 Human Male BAC"
1..2222
/notes="assembly_fragment"
2323..3785
/notes="assembly_fragment"
3886..6753
/notes="assembly_fragment"
6854..10036
/notes="assembly_fragment"
10137..14537
/notes="assembly_fragment"
14638..21177
/notes="assembly_fragment"
21278..30240
/notes="assembly_fragment"
30341..40123
/notes="assembly_fragment"
40224..47999
/notes="assembly_fragment
clone end:SP6
vector side:left"
48100..58235
/notes="assembly_fragment"
58336..67041
/notes="assembly_fragment"
67142..75921
/notes="assembly_fragment"
76022..86527
/notes="assembly_fragment"
86628..101334
/notes="assembly_fragment"
101435..117630
/notes="assembly_fragment"
117731..138293
/notes="assembly_fragment"
138394..157841
/notes="assembly_fragment"
157942..182152
/notes="assembly_fragment
clone end:T7
vector side:right"

ORIGIN
Query Match 55.2%; Score 667; DB 2: Length 182152;
Best Local Similarity 99.9%; Pred.No. 7.3e-298; Indels 0; Gaps 0;
Matches 717; Conservative 0; Mismatches 1;

QY 436 GAGTACCTCATGGGCAAGTTCCTTTTCCGGGCGAGCTACACGGGCTGGGCGCAACTCC 495
DB 10036 GAGTACCTCATGGGCAAGTTCCTTTTCCGGGCGAGCTACACGGGCTGGGCGCAACTCC 9977
QY 496 GTGGGCGACACCTTTGCTCAAGCTGCTTGGTCAAGTCTGGGCGACCCCTCGCGG 555
DB 9976 GTGGGCGACACCTTTGCTCAAGCTGCTTGGTCAAGTCTGGGCGACCCCTCGCGG 9917
QY 556 CCTGGGCGAAGGACAACTACTGGATGCTCAACCCCAAGGAGTACACCTTGGCGGAC 615
DB 9916 CCTGGGCGAAGGACAACTACTGGATGCTCAACCCCAAGGAGTACACCTTGGCGGAC 9857
QY 616 GGGGTCTTCCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGCGTGTCCCGCGCGCGG 675
DB 9856 GGGGTCTTCCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGCGTGTCCCGCGCGCGG 9797
QY 676 CTGGGCGCGGAGAGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 735
DB 9796 CTGGGCGCGGAGAGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 9737
QY 736 GCTTCGCGCGCGATGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 795
DB 9736 GCTTCGCGCGCGATGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 9677
QY 796 TTCTCCAGCTCCTTCGCTTCGACACCTCTCGGAGCGCTTCCGAGCGCTGCGCCCTC 855
```

```
DB 9676 TTCTCCAGCTCCTTCGCCATCGACAGCATCCTGCGCAAGCCCTTCGCGAGCGCGCGCTC 9617
QY 856 AGGACACGCGCGCGCGGAGAGCTTCAGTGGGCGCGCGCGCGCGCGCGCGCGCGCTGCC 915
DB 9616 AGGACACGCGCGCGCGGAGAGCTTCAGTGGGCGCGCGCGCGCGCGCGCGCGCGCTGCC 9557
QY 916 GCGTTCCCGCGCTCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGCGCGTAC 975
DB 9556 GCGTTCCCGCGCTCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGCGCGTAC 9497
QY 976 GCGCGCGCGAGCGCGCGCGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1035
DB 9496 GCGCGCGCGAGCGCGCGCGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9437
QY 1036 CCCCTCCTGCTTGCACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1095
DB 9436 CCCCTCCTGCTTGCACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9377
QY 1096 GCGCGCGCGCGCGCGCGCGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1153
DB 9376 GCGCGCGCGCGCGCGCGCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 9319

RESULT 6
AC012350/3
LOCUS Homo sapiens clone RP11-16N9, WORKING DRAFT SEQUENCE, 16 unordered
DEFINITION pieces.
ACCESSION AC012350
VERSION AC012350.3 GI:7381803
KEYWORDS HTG: HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-16N9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 178168)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukigaiter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Feerleira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL
COMMENT Submitted (25-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 1, 2000 this sequence version replaced gi:6479001.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://fcp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3593
Center clone name: 16_N9
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
```

Consensus quality: 172070 bases at least Q40  
Consensus quality: 174479 bases at least Q30  
Consensus quality: 175597 bases at least Q20  
Insert size: 170000; agarose-ff  
Insert size: 176668; sum-of-contigs  
Quality coverage: 6.4 in Q20 bases; agarose-ff  
Quality coverage: 6.2 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 16 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1694: contig of 1694 bp in length  
\* 1695  
\* 1794: gap of 100 bp  
\* 4437: contig of 2643 bp in length  
\* 4438  
\* 4537: gap of 100 bp  
\* 4538  
\* 6644: contig of 2107 bp in length  
\* 6645  
\* 6744: gap of 100 bp  
\* 6745  
\* 10092: contig of 3348 bp in length  
\* 10093  
\* 10192: gap of 100 bp  
\* 10193  
\* 11801: contig of 1609 bp in length  
\* 11802  
\* 11901: gap of 100 bp  
\* 11902  
\* 15584: contig of 3683 bp in length  
\* 15585  
\* 15684: gap of 100 bp  
\* 15685  
\* 19675: contig of 3991 bp in length  
\* 19676  
\* 19775: gap of 100 bp  
\* 19776  
\* 27913: contig of 8138 bp in length  
\* 27914  
\* 28013: gap of 100 bp  
\* 28014  
\* 34156: contig of 6143 bp in length  
\* 34157  
\* 34256: gap of 100 bp  
\* 34257  
\* 43325: contig of 9069 bp in length  
\* 43326  
\* 43425: gap of 100 bp  
\* 43426  
\* 55265: contig of 11840 bp in length  
\* 55266  
\* 55365: gap of 100 bp  
\* 55366  
\* 68308: contig of 12941 bp in length  
\* 68309  
\* 68408: gap of 100 bp  
\* 68409  
\* 82910: contig of 14504 bp in length  
\* 82911  
\* 83010: gap of 100 bp  
\* 83011  
\* 102181: contig of 19171 bp in length  
\* 102182  
\* 128058: contig of 100 bp  
\* 128059  
\* 128158: gap of 100 bp  
\* 128159  
\* 178168: contig of 50010 bp in length.

FEATURES

Location/Qualifiers  
1..178168  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="RP11-1639"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..1694  
/note="assembly\_fragment"  
1795..4437  
/note="assembly\_fragment"  
4538..6644  
/note="assembly\_fragment"  
6745..10092  
/note="assembly\_fragment"  
10193..11801  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
11902..15584  
/note="assembly\_fragment"  
15685..19675  
/note="assembly\_fragment"  
19776..27913  
/note="assembly\_fragment"  
28014..34156

misc_feature	/note="assembly_fragment" 34257..43325									
misc_feature	/note="assembly_fragment" 43426..55265									
misc_feature	/note="assembly_fragment" 55366..68306									
misc_feature	/note="assembly_fragment" 68407..82910									
misc_feature	/note="assembly_fragment" 83011..102181									
misc_feature	/note="assembly_fragment" clone_end:T7 vector_side:right 102282..128058									
misc_feature	/note="assembly_fragment" 128159..178168									
misc_feature	/note="assembly_fragment"									
ORIGIN										
Query Match 45.8%; Score 554; DB 2; Length 178168; Best Local Similarity 99.8%; Pred. No. 1.5e-245; Matches 604; Conservative 0; Mismatches 1; Indels 0; Gaps 0;										
QY	549	CTCGCGGCCCTGGGGCAAGGACAACTACTGATGCTCAACCCCAACAGCGAGTACACCTT	608							
DB	15584	CTCGCGGCCCTGGGGCAAGGACAACTACTGATGCTCAACCCCAACAGCGAGTACACCTT	15525							
QY	609	CGCCGACGGGTCTTCGCCCGCCGCGGCTCAGCCACCGCGCGCGCGCTCCCGCGC	668							
DB	15524	CGCCGACGGGTCTTCGCCCGCCGCGGCTCAGCCACCGCGCGCGCGCTCCCGCGC	15465							
QY	669	GCCCGGGTGGCGGCCCGAGGAGGCCCGGGCTCTCCCGCGCGCGCGCGCGCGCGCGC	728							
DB	15464	GCCCGGGTGGCGGCCCGAGGAGGCCCGGGCTCTCCCGCGCGCGCGCGCGCGCGCGC	15405							
QY	729	CGCCCGGCTCGCCCGCGCATGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	788							
DB	15404	CGCCCGGCTCGCCCGCGCATGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	15345							
QY	789	GCGCAAGTTCTCCAGCTCTTTGCCATCGACAGCATCTCTGGCGAAGCCCTTCGCGAGCG	848							
DB	15344	GCGCAAGTTCTCCAGCTCTTTGCCATCGACAGCATCTCTGGCGAAGCCCTTCGCGAGCG	15285							
QY	849	TGCGCTCAGGACACGCGCGCGCGCGCGCGCTTCAGTGGGGCGCGCGCGCTTCGCCGCG	908							
DB	15284	CGCGCTCAGGACACGCGCGCGCGCGCGCGCTTCAGTGGGGCGCGCGCGCTTCGCCGCG	15225							
QY	909	GCTGCGCGGTTCGCCGCGCTCTCTCCCGCGCGCGCTTCGACAGGGCGCTGTGCGCGTCTG	968							
DB	15224	GCTGCGCGGTTCGCCGCGCTCTCTCCCGCGCGCGCTTCGACAGGGCGCTGTGCGCGTCTG	15165							
QY	969	CGCGTACGCGCGCGCGCGCGCGCTGGGCGCGCGCGCGCGCGAGTGCACCGAC	1028							
DB	15164	CGCGTACGCGCGCGCGCGCGCGCTGGGCGCGCGCGCGCGAGTGCACCGAC	15105							
QY	1029	CGCGCGCGCCCTCTCTGTTCACCTCTCTCCCGCGCGCGCGCGCGCGCAAGCACTCCGAGG	1088							
DB	15104	CGCGCGCGCCCTCTCTGTTCACCTCTCTCCCGCGCGCGCGCGCGCGCAAGCACTCCGAGG	15045							
QY	1089	CCCGCGCGCGCGCGCGCGCACTGTATCTGCCCTTCGGGCTGCCCGCAGCCCTTCGAGGC	1148							
DB	15044	CCCGCGCGCGCGCGCGCGCACTGTATCTGCCCTTCGGGCTGCCCGCAGCCCTTCGAGGC	14985							
QY	1149	GGCCT 1153								
DB	14984	GGCCT 14980								

RESULT 7  
AC016269/c AC016269 74969 bp DNA linear HTG 13-JUL-2000  
LOCUS Homo sapiens clone RP11-1639, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC016269  
ACCESSION AC016269

```

VERSION      AC016269.3  GI:9123861
KEYWORDS     HTG; HTGS PHASE0.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 74969)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 74969)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.

-----
TITLE       Direct Submission
JOURNAL
COMMENT     Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 13, 2000 this sequence version replaced gi:7144942.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -----
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence.submissions@genome.wi.mit.edu
            -----
            Project Information
            Center Project name: L3560
            Center Clone name: 16_J_9
            -----
            * NOTE: This record contains 73 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1 944: contig of 944 bp in length
            * 945 1044: gap of 100 bp
            * 1045 1991: contig of 947 bp in length
            * 1992 2091: gap of 100 bp
            * 2092 3007: contig of 916 bp in length
            * 3008 3107: gap of 100 bp
            * 3108 4037: contig of 930 bp in length
            * 4038 4137: gap of 100 bp
            * 4138 5056: contig of 919 bp in length
            * 5057 5156: gap of 100 bp
            * 5157 6079: contig of 923 bp in length
            * 6080 6179: gap of 100 bp
            * 6180 7103: contig of 924 bp in length
            * 7104 7203: gap of 100 bp
            * 7204 8136: contig of 933 bp in length
            * 8137 8236: gap of 100 bp
            * 8237 9166: contig of 930 bp in length
            * 9167 9266: gap of 100 bp
            * 9267 10167: contig of 901 bp in length
            * 10168 10267: gap of 100 bp
            * 10268 11212: contig of 945 bp in length
            *
            * 944: contig of 944 bp in length
            * 945 1044: gap of 100 bp
            * 1045 1991: contig of 947 bp in length
            * 1992 2091: gap of 100 bp
            * 2092 3007: contig of 916 bp in length
            * 3008 3107: gap of 100 bp
            * 3108 4037: contig of 930 bp in length
            * 4038 4137: gap of 100 bp
            * 4138 5056: contig of 919 bp in length
            * 5057 5156: gap of 100 bp
            * 5157 6079: contig of 923 bp in length
            * 6080 6179: gap of 100 bp
            * 6180 7103: contig of 924 bp in length
            * 7104 7203: gap of 100 bp
            * 7204 8136: contig of 933 bp in length
            * 8137 8236: gap of 100 bp
            * 8237 9166: contig of 930 bp in length
            * 9167 9266: gap of 100 bp
            * 9267 10167: contig of 901 bp in length
            * 10168 10267: gap of 100 bp
            * 10268 11212: contig of 945 bp in length
            *
            * 11213 11312: gap of 100 bp
            * 11313 12250: contig of 938 bp in length
            * 12251 12350: gap of 100 bp
            * 12351 13277: contig of 927 bp in length
            * 13278 13377: gap of 100 bp
            * 13378 14275: contig of 898 bp in length
            * 14276 14376: gap of 100 bp
            * 14377 15281: contig of 906 bp in length
            * 15282 15382: gap of 100 bp
            * 15383 16289: contig of 918 bp in length
            * 16290 16390: gap of 100 bp
            * 16391 17341: contig of 942 bp in length
            * 17342 17442: gap of 100 bp
            * 17443 18359: contig of 918 bp in length
            * 18360 18459: gap of 100 bp
            * 18460 19396: contig of 937 bp in length
            * 19397 19498: gap of 100 bp
            * 19499 20457: contig of 961 bp in length
            * 20458 20558: gap of 100 bp
            * 20559 21482: contig of 925 bp in length
            * 21483 22515: contig of 933 bp in length
            * 22516 23514: contig of 899 bp in length
            * 23515 23614: gap of 100 bp
            * 23615 24531: contig of 916 bp in length
            * 24532 24630: gap of 100 bp
            * 24631 25553: contig of 923 bp in length
            * 25554 25653: gap of 100 bp
            * 25654 26595: contig of 942 bp in length
            * 26596 26695: gap of 100 bp
            * 26696 27627: contig of 932 bp in length
            * 27628 27727: gap of 100 bp
            * 27728 28660: contig of 933 bp in length
            * 28661 29792: contig of 932 bp in length
            * 29793 30708: contig of 916 bp in length
            * 30709 30808: gap of 100 bp
            * 30809 31728: contig of 920 bp in length
            * 31729 31828: gap of 100 bp
            * 31829 32740: contig of 912 bp in length
            * 32741 32840: gap of 100 bp
            * 32841 33761: contig of 921 bp in length
            * 33762 33861: gap of 100 bp
            * 33862 34804: contig of 943 bp in length
            * 34805 34904: gap of 100 bp
            * 34905 35835: contig of 931 bp in length
            * 35836 35935: gap of 100 bp
            * 35936 36882: contig of 947 bp in length
            * 36883 37911: contig of 929 bp in length
            * 37912 38012: gap of 100 bp
            * 38013 38936: contig of 925 bp in length
            * 38937 39036: gap of 100 bp
            * 39037 39990: contig of 954 bp in length
            * 39991 40090: gap of 100 bp
            * 40091 41004: contig of 914 bp in length
            * 41005 42018: contig of 914 bp in length
            * 42019 43015: gap of 100 bp
            * 43016 43115: contig of 897 bp in length
            * 43116 44052: contig of 937 bp in length
            * 44053 44152: gap of 100 bp
            * 44153 45059: contig of 907 bp in length
            * 45060 46092: contig of 933 bp in length
            * 46093 46192: gap of 100 bp
            * 46193 47110: contig of 918 bp in length
            * 47111 47210: gap of 100 bp
            * 47211 48138: contig of 928 bp in length
            * 48139 48238: gap of 100 bp

```





KEYWORDS	Mus musculus (house mouse)	REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
SOURCE	Mus musculus	AUTHORS	1 (bases 1 to 2046)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1513)		
AUTHORS	Hong,H.K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S., Justice,M.J. and Chakravarti,A.		
TITLE	The winged helix/forhead transcription factor Foxq1 regulates differentiation of hair in satin mice		
JOURNAL	Genesis 29 (4), 163-171 (2001)		
MEDLINE	21207067		
PUBMED	11309849		
REFERENCE	2 (bases 1 to 1513)		
AUTHORS	Hong,H.-K., Noveroske,J.K., Justice,M.J. and Chakravarti,A.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-MAY-1999) Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA		
FEATURES	Location/Qualifiers		
source	1..1513		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/strain="129"		
	/db_xref="taxon:10090"		
gene	<158..>1360		
	/gene="Hfh1"		
mRNA	<158..>1360		
	/gene="Hfh1"		
	/product="Hfh1"		
CDS	158..1360		
	/gene="Hfh1"		
	/codon_start=1		
	/product="Hfh1"		
	/protein_id="AA74524.1"		
	/db_xref="GI:8347145"		
	/translations="MKLEVFVRAAHGDKWGLSAGSDVPSPLAAGDSDLSGDG DCAANGPAAGSGADLEGGERNSGSPSAQGPATDSRTQSAAGCAGGVGGG EGARSKPTRRPFPYSYALIAIMARDGAGRLTAELNYLMGFPPFRGSGYTGWR NSVRLHSLNDCVFLKLRDPSRWGKDYMLNPNEYTFADGVFRRRKLRSHRTTV SASGLRPEEPAPPGTQPPAPARSPIARSPAROEBSRSPASFSFADISLTK PFRSRDGDALGVQLPWGAPCPPIRAPYALIPAAAGGALLPLCAVGASEPTLLASR GTVPQAPAPLAPLSTAAAPAKPFRFPETAGAAHLYCPLRLPTALQAAAACGGPHLS YPVETLLA"		
ORIGIN			
Query Match	6.48; Score 77; DB 10; Length 1513;		
Best Local Similarity	99.28; Pred. No. 1.9e-24;		
Matches 127; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
Qy	526 TTGCTCAAGTGTGTCGCGACCCCTCGCGCCCTGGGGCAAGGACAACTACTGGATGCTC 585		
Db	674 TTGCTCAAGTGTGTCGCGACCCCTCGCGCCCTGGGGCAAGGACAACTACTGGATGCTC 733		
Qy	586 AACCCCAACGAGTAGTACACTTCGCCGACGGGTCTTCGCCGCCGCCGCGCAGGCCCTC 645		
Db	734 AACCCCAACGAGTAGTACACTTCGCCGACGGGTCTTCGCCGCCGCCGCGCAGGCCCTC 793		
Qy	646 AGCCACCG 653		
Db	794 AGCCACCG 801		
RESULT 10			
BC047155			
LOCUS	BC047155 2046 bp mRNA linear ROD 07-OCT-2003		
DEFINITION	Mus musculus forkhead box Q1, mRNA (cdna clone MGC:54661 IMAGE:6489152), complete cds.		
ACCESSION	BC047155		
VERSION	BC047155.1 GI:28502981		
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

FEATURES	source		
1..2046	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="FVB/N-3"		
	/db_xref="taxon:10090"		
	/clone="MGC:54661 IMAGE:6489152"		
	/tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."		
	/clone_lib="NCI CGAP_Mam2"		
	/lab_host="DH10B"		
	/note="Vector: pCWV-SPORT6"		
1..2046	/gene="Foxq1"		
	/note="synonym: HPH-1"		
	/db_xref="LocusID:15220"		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: ISAK Plate: 100 Row: n Column: 18  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6754191.

Location/Qualifiers  
 1..2046

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N-3"

/db\_xref="taxon:10090"

/clone="MGC:54661 IMAGE:6489152"

/tissue\_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."

/clone\_lib="NCI CGAP\_Mam2"

/lab\_host="DH10B"

/note="Vector: pCWV-SPORT6"

1..2046

/gene="Foxq1"

/note="synonym: HPH-1"

/db\_xref="LocusID:15220"

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2046)  
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,I., Max,S.I., Wang,J., Haieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fanev,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,B.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932  
 2 (bases 1 to 2046)  
 Strausberg,R.  
 Direct Submission  
 Submitted (21-FEB-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [gcgaps-remail.nih.gov](mailto:gcgaps-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland.  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc\\_mgc@nri.nih.gov](mailto:nisc_mgc@nri.nih.gov)  
 Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Statrippop,S., Thomas,P.J., Touchman,J.W., Tsurgueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,J.-H. and Green,E.D.

```
/db_xref="MGI:1298228"
481..1683
/codon_start=1
/product="forkhead box Q1"
/protein_id="AAH47155.1"
/db_xref="GI:28502982"
/db_xref="LocusID:15220"
/translations="MKLEVFVRAAHGDKMGSDLEAGSSDVPSPLSAAGDSDLSGSDG
DCANSPAGSGAGDLEGGGERNSSGGPSAQDPEATDSDRTQASAGPCAGGVGGG
EGARSKPYTRRPKPPYSYIALIAMAIRDSAGRLTLAEINELMKGPPFRGSGYTGWR
NSVRHNLSDNCFVKLRDPSRPMGKNDYMLNPNSYTFADGVFRERKRLSHRTTV
SASGLPEAPPGPAGTPOAPARSSPIARSPARQERSSPASKFSSFAIDSLISK
PFRSRDGSALGVOLPWGAAPCPPLRAYPALLPAAPGALLPLCAYGASEPTLLASR
GTEVQPAAPLLAPLSTAPAKPFRGPETAGAAHLYCPLRLPTALQAAAACGPGPHLS
YFVETLLA"
misc_feature
817..1089
/notes="FH; Region: FORKHEAD"
/db_xref="CDD:smart00339"

ORIGIN
Query Match 6.4%; Score 77; DB 10; Length 2046;
Best Local Similarity 99.2%; Pred. No. 1.8e-24;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 526 TTGCTCAAGTGTCGCGGACCCCTCGCGCCCTCGGGCAAGGACAACTACTGGATGCTC 585
DB 997 TTGCTCAAGTGTCGCGGACCCCTCGCGCCCTCGGGCAAGGACAACTACTGGATGCTC 1056

QY 586 AACCCCAACAGGAGTACACCTTCGCGGAGGGGTCTTCGCGCGCGGCAAGGCGCTC 645
DB 1057 AACCCCAACAGGAGTACACCTTCGCGGAGGGGTCTTCGCGCGCGGCAAGGCGCTC 1116

QY 646 AGCCACCG 653
DB 1117 AGCCACCG 1124

RESULT 11
AF010405
LOCUS AF010405.2 GI:6716869
DEFINITION Mus musculus fork head transcription factor (Hfh-1L) gene, complete cds.
ACCESSION AF010405
VERSION AF010405.2
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4763)
TITLE Mouse HNF-3/fork head homolog-1-like gene: structure, chromosomal
JOURNAL location, and expression in adult and embryonic kidney
MEDLINE DNA Cell Biol. 17 (8), 679-688 (1998)
PUBMED 98392851
REFERENCE 9726250
2 (bases 1 to 4763)
AUTHORS Frank, S. and Zoll, B.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1997) University of Goettingen, Institute of
Human Genetics, Gosslerstr 12d, Goettingen 37073, Germany
REFERENCE 3 (bases 1 to 4763)
AUTHORS Pasche, B., Bieller, A. and Zoll, B.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2000) University of Goettingen, Institute of
Human Genetics, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany
REMARK Sequence update by submitter
COMMENT On Jan 20, 2000 this sequence version replaced gi:2997586.
FEATURES
source
1..4763
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129"
```

```
/db_xref="taxon:10090"
/chromosome="13"
/tissue_type="kidney"
<2467..>3669
/gene="Hfh-1L"
/notes="HNF-3/fork head homologue-1 like"
<2467..>3669
/gene="Hfh-1L"
/product="fork head transcription factor"
2467..3669
/notes="Hfh-1L"
/notes="member of HNF-3/fork head transcription factor family"
/codon_start=1
/product="fork head transcription factor"
/protein_id="AAC12973.2"
/db_xref="GI:6716870"
/translation="MKLEVFVRAAHGDKMGSDLEAGSSDVPSPLSAAGDSDLSGSDG
DCANSPAGSGAGDLEGGGERNSSGGPSAQDPEATDSDRTQASAGPCAGGVGGG
EGARSKPYTRRPKPPYSYIALIAMAIRDSAGRLTLAEINELMKGPPFRGSGYTGWR
NSVRHNLSDNCFVKLRDPSRPMGKNDYMLNPNSYTFADGVFRERKRLSHRTTV
SASGLPEAPPGPAGTPOAPARSSPIARSPARQERSSPASKFSSFAIDSLISK
PFRSRDGSALGVOLPWGAAPCPPLRAYPALLPAAPGALLPLCAYGASEPTLLASR
GTEVQPAAPLLAPLSTAPAKPFRGPETAGAAHLYCPLRLPTALQAAAACGPGPHLS
YFVETLLA"
misc_feature
2782..3123
/gene="Hfh-1L"
/notes="encodes DNA binding domain"

ORIGIN
Query Match 6.4%; Score 77; DB 10; Length 4763;
Best Local Similarity 99.2%; Pred. No. 1.6e-24;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 526 TTGCTCAAGTGTCGCGGACCCCTCGCGCCCTCGGGCAAGGACAACTACTGGATGCTC 585
DB 2983 TTGCTCAAGTGTCGCGGACCCCTCGCGCCCTCGGGCAAGGACAACTACTGGATGCTC 3042

QY 586 AACCCCAACAGGAGTACACCTTCGCGGAGGGGTCTTCGCGCGCGGCAAGGCGCTC 645
DB 3043 AACCCCAACAGGAGTACACCTTCGCGGAGGGGTCTTCGCGCGCGGCAAGGCGCTC 3102

QY 646 AGCCACCG 653
DB 3103 AGCCACCG 3110

RESULT 12
AL589738
LOCUS AL589738 187709 bp DNA linear ROD 29-JUN-2002
DEFINITION Mouse DNA sequence from clone RP23-322U11 on chromosome 13, complete sequence.
ACCESSION AL589738
VERSION AL589738.6 GI:15422196
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 187709)
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
COMMENT On Sep 3, 2001 this sequence version replaced gi:13445573.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-322J11 is from the RPECI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBacE3.6

----- Genome Center  
Center: UK Medical Research Council  
Center code: UK-MRC  
Web site: <http://mrcseq.har.mrc.ac.uk>  
Contact: [mouse@har.mrc.ac.uk](mailto:mouse@har.mrc.ac.uk)

-----

FEATURES  
source  
1. 197709 Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosomes="13"  
/clone="RP23-322J11"  
/clone\_lib="RPECI-23"

ORIGIN  
Query Match 6.4%; Score 77; DB 10; Length 187709;  
Best Local Similarity 99.2%; Pred. No. 1.1e-24;  
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 526 TTGCTCAAGTGTGCGACCCCTCGCGCCCTGCGGCAAGCAACACTACTGGATGTC 585  
Db 136222 TTGCTCAAGTGTGCGACCCCTCGCGCCCTGCGGCAAGCAACACTACTGGATGTC 136281

QY 586 AACCCCAACAGCAGTACACTTCGCGACGGGTTCTCCGCGCGCGCAAGCGCCTC 645  
Db 136282 AACCCCAACAGCAGTACACTTCGCGACGGGTTCTCCGCGCGCGCAAGCGCCTC 136341

QY 646 AGCCACCG 653  
Db 136342 AGCCACCG 136349

RESULT 13  
AF153193  
LOCUS AF153193 1576 bp DNA linear ROD 24-AUG-2001  
DEFINITION Rattus rattus winged helix/forkhead transcription factor HFH1 (Hfh1) gene, complete cds.  
ACCESSION AF153193  
VERSION AF153193.1 GI:8132282  
KEYWORDS Rattus rattus (black rat)  
SOURCE Rattus rattus  
ORGANISM Rattus rattus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 1576)  
AUTHORS Hong H.-K., Noveroske J.K., Headon D.J., Liu T., Sy M.S., Justice M.J. and Chakravarti A.  
TITLE The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice  
JOURNAL Genesis 29 (4), 163-171 (2001)  
MEDLINE 21207067  
PUBMED 11309849  
REFERENCE 2 (bases 1 to 1576)  
AUTHORS Hong H.-K., Noveroske J.K., Justice M.J. and Chakravarti A.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-1999) Genetics, Case Western Reserve University,

10900 Euclid Ave., Cleveland, OH 44106-4955  
Location/Qualifiers  
1. 1576  
/organism="Rattus rattus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10117"  
<95> .>1297  
/genes="Hfh1"  
<95> .>1297  
/genes="Hfh1"  
/product="winged helix/forkhead transcription factor HFH1"  
95. 1297  
/genes="Hfh1"  
/codon\_start=1  
/product="winged helix/forkhead transcription factor HFH1"  
/protein\_id="AA573234.1"  
/db\_xref="GI:8132283"  
/translation="MKLEVPAPRAAHGDMGSLGAGSDVSPISAAAGDLSGSDG  
DCAANGPAAGRGAVDLEGGGERSGGASTQDDPEVTGSRTPQSFVPGAGSVGG  
EGARSKPYTRRPKPYSYIALIAIRDSAGRLTLAINEYLMGKFPFRGYSYQWR  
NSVRHNLSLNDFVKVLRDPSRPWGNKDNMLNPNSEYTPADGVFRRRLSHRTTV  
SASGLRPEEAPPGAGTPOAPTAGSPAROBEGSSPASKTSFSSPAIDSLISK  
PFRSRDGDPAQLVQLPWSPAACPPIRAYPALIPASSGGLPLCAYGAGEPTLLASR  
GREVQPAPLLPLAPLSTAAAPAKFRFETAGAAHLICPLRLPTALQAARACGPGPHLS  
YRVETLLA"

ORIGIN  
Query Match 4.4%; Score 53; DB 10; Length 1576;  
Best Local Similarity 100.0%; Pred. No. 2.4e-13;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 TACACCTTCGCGACGGGGTCTTCGCGCGCGCGCGAGCGCTCAGCCACCG 653  
Db 686 TACACCTTCGCGACGGGGTCTTCGCGCGCGCGCGAGCGCTCAGCCACCG 738

RESULT 14  
RATHFH1  
LOCUS RATHFH1 1760 bp mRNA linear ROD 22-AUG-1995  
DEFINITION Rattus norvegicus HNF-3/forkhead homolog-1 (HFH-1) mRNA, complete cds.  
ACCESSION L13201  
VERSION L13201.1 GI:951067  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 1760)  
AUTHORS Clevidence D.E., Overdier D.G., Tao W., Qian X., Pani L., Lai E. and Costa R.H.  
TITLE Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (9), 3948-3952 (1993)  
MEDLINE 93248207  
PUBMED 7683413  
COMMENT On Aug 22, 1995 this sequence version replaced gi:550512.  
Original source text: rat.  
Location/Qualifiers  
1. 1760  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/sex="male"  
/tissue\_type="lung"  
/dev\_stage="9 week old"  
/tissue\_lib="Clontech"  
1. 1760  
/gene="HFH-1"  
250. 1569  
/gene="HFH-1"

FEATURES  
source  
1. 1760  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/sex="male"  
/tissue\_type="lung"  
/dev\_stage="9 week old"  
/tissue\_lib="Clontech"  
1. 1760  
/gene="HFH-1"  
250. 1569  
/gene="HFH-1"

/note="bp 84. .889 winged-helix DNA binding domain"  
/codon\_start=1  
/product="HNF-3/forhead homolog-1"  
/protein\_id="AAA74561.1"  
/db\_xref="GI:550513"  
/translation="MGSLEGAGSSDVPSPLSAAGDDSLGSDGCAANSPPAAGSGAD  
LEGGGERNSSGASGTDQDDPVTGRTQAGVPGPCAGSVGGEGARSKPVTTRPKPP  
YSYIALIMARSDAGSLTLAEINEVLMGKFPFPGSYTGWRNSVRHNLNDCFVK  
VLRPSRWGKNTYMLNPNSEYTPADGVFRRKRRLSHRTVSASGYGGSPGPAG  
TPQAPTAGSPISPARSPAGSPSPSSFAIDSIILSKPFRSRDGTGLRGCSY  
FGALPAPRCAPISFPFRPAVPCRSVLTRAAAGAVARGAARGAPVARGAPUH  
RGPSPAIRSGDRRGRLVLPFTAAGPAGRGRLSGAVFLPGDAASTLGAWSSAG  
GSRVELGAPFVMEGTFCGHWAGSRAVHFPPSPRNRDFFSTLHQM"

ORIGIN

Query Match 4.4%; Score 53; DB 10; Length 1760;  
Best Local Similarity 100.0%; Pred. No. 2.4e-13;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 TACACTTCGCGAGCGGGTCTTCGCGCGCCGCGCCGCGAGCGCTCAGCCACCG 653  
DB 796 TACACTTCGCGAGCGGGTCTTCGCGCGCCGCGCCGCGAGCGCTCAGCCACCG 848

RESULT 15

AC119497 195184 bp DNA linear HTG 20-NOV-2002  
Rattus norvegicus clone CH230-521P8, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 4 unordered pieces.  
AC119497  
AC119497.5 GI:25137938  
HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE

1 (bases 1 to 195184)  
Muzny, D.Marie, Metzker, M.Lee, Abruzzo, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kowar, C.,  
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuewa, L., Louleghed, H., Lozano, R.J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, B.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Popper, F., Poidexter, A., Popovic, D., Primus, E., Pu, L.-L.,  
Pua, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmari, K.,  
Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 195184)  
Worley, K.C.  
Direct Submission  
Submitted (27-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 195184)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 20, 2002 this sequence version replaced gi:23195197.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GWQI  
Center clone name: CH230-521P8  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 158913 bases at least Q40  
Consensus quality: 162109 bases at least Q30  
Consensus quality: 164133 bases at least Q20  
Estimated insert size: 161675; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 4 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
\* 1 16506: contig of 16506 bp in length  
\* 16507 18606: gap of unknown length  
\* 16607 180638: contig of 164032 bp in length  
\* 180639 180738: gap of unknown length  
\* 180739 182447: contig of 1709 bp in length  
\* 182448 182547: gap of unknown length

```
FEATURES * 182548 195184: contig of 12637 bp in length.
Source      Location/Qualifiers
1..195184
/organism="Rattus norvegicus"
/mol_type="Genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-521F8"
465..1181
/note="clone boundary"
clone_end:T7
site:
end_sequence:BZ266854"
misc_feature 1768..2639
/note="clone boundary"
clone_end:Sp6
site:
end_sequence:BZ267695"
misc_feature 16507..17640
/note="wgs_end_extension"
clone_end:Sp6"
misc_feature 71528..72576
/note="wgs_end_extension"
clone_end:Sp6"
misc_feature 76593..78955
/note="wgs_end_extension"
clone_end:Sp6"
misc_feature 86077..87329
/note="wgs_end_extension"
clone_end:Sp6"
misc_feature 106805..108354
/note="wgs_end_extension"
clone_end:Sp6"
misc_feature 129622..130958
/note="wgs_end_extension"
clone_end:Sp6"
```

ORIGIN

```
Query Match      4.4%; Score 53; DB 2; Length 195184;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 53, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 601 TACACCTTCGCCGACGGGGTCTTCGGCGCGCGCGCGCGCGCTCAGCCACCG 653
      |||||
Db 136744 TACACCTTCGCCGACGGGGTCTTCGGCGCGCGCGCGCGCGCTCAGCCACCG 136796
```

Search completed: March 31, 2004, 03:52:56  
Job time : 4683 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 01:11:33 ; Search time 504 Seconds  
(without alignments)  
10190.618 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggagggttgcttgcgt.....tggagacgctcctagcttga 1209

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001s:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002s:\*
- 7: Geneseqn2003s:\*
- 8: Geneseqn2003bs:\*
- 9: Geneseqn2003cs:\*
- 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1209	100.0	1209	6	ABQ81557 Gene up-r
2	53	4.4	1760	7	ABT42112 Toxicity
3	37	3.1	658	6	ABQ39023 Oligonucle
4	37	3.1	658	6	ABQ39023 Oligonucle
5	30	2.5	1387	5	ABQ72684 DNA encod
6	30	2.5	3786	5	ABQ72683 DNA encod
7	27	2.2	514	6	ABQ34097 Oligonucle
8	27	2.2	514	6	ABQ34096 Oligonucle
9	27	2.2	741	6	ABQ34746 Oligonucle
10	27	2.2	741	6	ABQ34747 Oligonucle
11	27	2.2	3482	7	ABX76288 Lung canc
12	26	2.2	738	6	ABQ25892 Oligonucle
13	26	2.2	738	6	ABQ25893 Oligonucle
14	25	2.1	441	6	ABT80120 Human ova
15	25	2.1	890	6	ABQ34485 Oligonucle
16	25	2.1	890	6	ABQ34484 Oligonucle
17	25	2.1	1965	9	ABQ31305 Testoster
18	25	2.1	2187	7	ABQ34783 Coding se
19	25	2.1	4895	2	AAQ79378 Human N-m
20	25	2.1	4895	2	AAQ79377 Human N-m
21	25	2.1	4895	3	AAQ38731 Human NMD
22	25	2.1	4895	3	AAQ38731 Human NMD
23	25	2.1	4895	6	AB199196 Human NMD

24	25	2.1	4695	6	AAL47384	AA147384 Human NMD
25	25	2.1	4695	7	ABX98543	ABX98543 Human N-m
26	25	2.1	4695	7	ABX92892	ABX92892 Human N-m
27	25	2.1	4695	7	ABX77688	ABX77688 DNA encod
28	25	2.1	4695	8	ACD98353	ACD98353 Human NMD
29	25	2.1	4695	9	AAD59685	AAD59685 Human NMD
30	24	2.0	51	4	AAL28440	AA128440 Human SNP
31	24	2.0	146	3	AAC09019	AAC09019 Human sec
32	24	2.0	159	4	AAH27142	AAH27142 Human cyc
33	24	2.0	310	6	AAH37498	AAH27142 COX2 prom
34	24	2.0	429	3	AAC57690	AAH37498 COX2 prom
35	24	2.0	1056	6	ABZ35308	ABZ35308 Arachidon
36	24	2.0	1782	8	ACC85255	ABZ35308 Human gen
37	24	2.0	1860	2	AAQ50632	ACC85255 Craniofac
38	24	2.0	1979	4	ABL58042	AAQ50632 Brain fac
39	24	2.0	2271	6	ABZ35021	ABL58042 Human cyc
40	24	2.0	2487	9	ADB75311	ABZ35021 Human gen
41	24	2.0	2563	7	ABX63477	ADB75311 Prostata
42	24	2.0	3387	2	AAQ71002	ABX63477 Human CDN
43	24	2.0	3387	2	AAQ89376	AAQ71002 Cyclooxyg
44	24	2.0	3387	3	AAQ34993	AAQ89376 Human cyc
45	24	2.0	3387	3	AAF21115	AAQ34993 Human ade
						AAF21115 Human_low

#### ALIGNMENTS

RESULT 1  
ABQ81557  
ID ABQ81557 standard; cDNA; 1209 BP.  
XX  
AC ABQ81557;  
XX  
DT 30-DEC-2002 (first entry)  
XX  
DE Gene up-regulated in metastatic colorectal cancer.  
XX  
KW Colorectal cancer; Metastasis; differential expression; cytostatic;  
KW diagnosis; gene therapy; vaccine; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1209  
FT /\*tag= a  
FT /product= "Metastatic colon cancer polypeptide"  
XX  
PN WO200268677-A2.  
XX  
PD 06-SEP-2002.

XX 27-FEB-2002; 2002WO-US006001.  
XX 27-FEB-2001; 2001US-0272206P.  
PR 02-APR-2001; 2001US-0281149P.  
PR 17-APR-2001; 2001US-0284555P.  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX Mack DH, Markowitz SD;  
XX WPI: 2002-698677/75.  
DR P-FSDB; ABP54691.

XX New genes that are up- or down-regulated in colorectal cancer, useful for  
PT diagnosing colorectal cancer in a subject, or for identifying modulators  
PT of colorectal cancer-associated proteins and genes for treating  
PT colorectal cancer.

XX Claim 5; Page 252; 260pp; English.

XX The present sequence is the nucleotide sequence of a human gene that



```

XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX WPI; 2003-148464/14.
XX
XX Predicting at least one toxic effect of a compound, useful for toxicity
XX PT modeling, comprises preparing a gene expression profile of a tissue or
XX PT cell sample exposed to the compound, and comparing the gene expression
XX PT profile to a database.
XX
XX Example 4; Page; 446pp; English.
XX
XX The invention relates to a novel method of predicting at least one toxic
XX effect of a compound. The method comprises a gene expression profile of a
XX tissue or cell sample exposed to the compound, and comparing the gene
XX expression profile to a database comprising at least part of the data or
XX information given in the specification. The methods are useful for
XX predicting at least one toxic effect of a compound, predicting the
XX progression of a toxic effect of a compound, predicting the renal
XX toxicity of a compound, or identifying toxicity markers in tissues or
XX cells exposed to known renal toxin. The genes are useful as toxicity
XX markers in drug screening and toxicity assays, in monitoring disease or
XX physiological states, or disease progression. This polynucleotide
XX represents a rat DNA sequence relating to the toxic effect database
XX described in the specification. NOTE: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from the World Intellectual Property
XX Organization
XX
XX Sequence 1760 BP; 301 A; 615 C; 535 G; 309 T; 0 U; 0 Other;
XX
XX Query Match 4.4%; Score 53; DB 7; Length 1760;
XX Best Local Similarity 100.0%; Pred. No. 1e-13;
XX Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 601 TACACCTTCGCGAGCGGGTCTTCGCGCGCGCGCGCGCTCAGCCACCG 653
XX DB 796 TACACCTTCGCGAGCGGGTCTTCGCGCGCGCGCGCGCTCAGCCACCG 848
XX
XX RESULT 3
XX ABQ39023
XX ID ABQ39023 standard; DNA; 658 BP.
XX
XX AC ABQ39023;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25614.
XX
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX
XX PP 01-SEP-2000; 2000DE-01043826.
XX
XX PR 05-SEP-2000; 2000DE-01044543.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX

```

```

XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX PT diagnosis and prognosis, comprises selective hybridization of amplicons
XX PT from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNP's); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ54121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention.
XX
XX Sequence 658 BP; 173 A; 310 C; 108 G; 67 T; 0 U; 0 Other;
XX
XX Query Match 3.1%; Score 37; DB 6; Length 658;
XX Best Local Similarity 100.0%; Pred. No. 1.6e-06;
XX Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 699 CTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735
XX DB 102 CTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 138
XX
XX RESULT 4
XX ABQ39022/C
XX ID ABQ39022 standard; DNA; 658 BP.
XX
XX AC ABQ39022;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25613.
XX
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX
XX PP 01-SEP-2000; 2000DE-01043826.
XX
XX PR 05-SEP-2000; 2000DE-01044543.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX PT diagnosis and prognosis, comprises selective hybridization of amplicons
XX PT from chemically treated DNA.

```



sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Query Match 2.5%; Score 30; DB 5; Length 3786;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 AAGCCCCCTACTCGTACATCGGCTCATC 381  
DB 322 AAGCCCCCTACTCGTACATCGGCTCATC 351

RESULT 7  
ABQ34097  
ID ABQ34097 standard; DNA; 514 BP.

AC ABQ34097;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20688.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; gastrointestinal; respiratory system; single nucleotide polymorphism; SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-RP010074.

PR 01-SEP-2000; 2000DE-01043826.

PS 05-SEP-2000; 2000DE-01044543.

PT (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert

cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention

Query Match 2.2%; Score 27; DB 6; Length 514;  
Best Local Similarity 100.0%; Pred. No. 0.049;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 CCCCCCTACTCGTACATCGGCTCATC 381

DB 138 CCCCCCTACTCGTACATCGGCTCATC 164

RESULT 8

ABQ34096/C

ID ABQ34096 standard; DNA; 514 BP.

AC ABQ34096;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20687.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; gastrointestinal; respiratory system; single nucleotide polymorphism; SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-RP010074.

PR 01-SEP-2000; 2000DE-01043826.

PS 05-SEP-2000; 2000DE-01044543.

PT (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member,

CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX Sequence 514 BP; 78 A; 42 C; 205 G; 189 T; 0 U; 0 Other;  
SQ  
Query Match 2.2%; Score 27; DB 6; Length 514;  
Best Local Similarity 100.0%; Pred. No. 0.049;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 355 CCCCCCTACTCGTACATCGCGCTCATC 381  
DB 377 CCCCCCTACTCGTACATCGCGCTCATC 351  
RESULT 9  
ABQ34746/C  
ID ABQ34746 standard; DNA; 741 BP.  
XX  
AC ABQ34746;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21337.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
FN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP010074.  
XX  
PR 01-SEP-2000; 2000DE-01043826.  
XX  
PR 05-SEP-2000; 2000DE-01044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of

CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX Sequence 741 BP; 112 A; 62 C; 259 G; 308 T; 0 U; 0 Other;  
SQ  
Query Match 2.2%; Score 27; DB 6; Length 741;  
Best Local Similarity 100.0%; Pred. No. 0.048;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 355 CCCCCCTACTCGTACATCGCGCTCATC 381  
DB 285 CCCCCCTACTCGTACATCGCGCTCATC 259  
RESULT 10  
ABQ34747  
ID ABQ34747 standard; DNA; 741 BP.  
XX  
AC ABQ34747;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21338.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
FN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP010074.  
XX  
PR 01-SEP-2000; 2000DE-01043826.  
XX  
PR 05-SEP-2000; 2000DE-01044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central

CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
 CC particularly by detecting mutations or single nucleotide polymorphisms  
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
 CC investigating cell differentiation. The method allows the methylation  
 CC status of many C residues to be determined simultaneously. ABQ13410-  
 CC ABQ054121 represent genomic DNA sequences used to illustrate the method  
 CC for determining the degree of cytosine methylation described in the  
 CC disclosure of the invention

XX SQ Sequence 741 BP; 308 A; 259 C; 62 G; 112 T; 0 U; 0 Other;  
 XX  
 XX Query Match 2.2%; Score 27; DB 6; Length 741;  
 XX Best Local Similarity 100.0%; Pred. No. 0.048; 0; Indels 0; Gaps 0;  
 XX Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 CCCCCCTACTCGTACATCGCGTCTATC 381  
 DB 457 CCCCCCTACTCGTACATCGCGTCTATC 483

RESULT 11  
 ABX76288  
 ID ABX76288 standard; DNA; 3482 BP.  
 XX AC ABX76288;  
 XX  
 XX 02-APR-2003 (first entry)  
 XX  
 XX Lung cancer-associated polynucleotide #152.  
 XX  
 XX Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;  
 XX antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
 XX small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
 XX chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
 XX interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
 XX  
 XX Unidentified.  
 XX  
 XX WO200286443-A2.  
 XX  
 XX 31-OCT-2002.  
 XX  
 XX 18-APR-2002; 2002WO-US012476.  
 XX  
 XX 18-APR-2001; 2001US-0284770P.  
 XX 10-MAY-2001; 2001US-0290492P.  
 XX 09-NOV-2001; 2001US-0339245P.  
 XX 13-NOV-2001; 2001US-0350666P.  
 XX 29-NOV-2001; 2001US-0334370P.  
 XX 12-APR-2002; 2002US-0372246P.  
 XX  
 XX (BOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 XX Aziz N, Murray R;  
 XX  
 XX WPI; 2003-093161/08.  
 XX P-PSDB; ABUS6559.  
 XX  
 XX Detecting a lung cancer-associated transcript in a cell from a patient  
 XX for treating lung cancer, by contacting a biological sample from the  
 XX patient with a polynucleotide that exhibits increased or decreased  
 XX expression in lung cancer.  
 XX  
 XX Claim 22; Page 303-304; 453pp; English.  
 XX  
 XX The invention relates to a method for detecting a lung cancer-associated  
 XX transcript in a cell from a patient, comprising contacting a biological  
 XX sample from the patient with a polynucleotide that selectively hybridises  
 XX to a sequence that is at least 80 % identical to a gene that exhibits  
 XX increased or decreased expression in lung cancer samples. Lung cancer-  
 XX associated polynucleotides and polypeptides are used for identifying a  
 XX compound that modulates a lung cancer-associated polypeptide, for  
 XX inhibiting proliferation of a lung cancer-associated cell to treat lung

CC cancer in a patient and for treating a mammal having lung cancer by  
 CC administering a modulatory compound identified. The methods are useful  
 CC for treating lung cancer, such as small cell lung cancer, non-small cell  
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful  
 CC for diagnostic purposes and as targets for screening for therapeutic  
 CC compounds that modulate lung cancer, such as antibodies. Sequences  
 CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the  
 CC invention

XX SQ Sequence 3482 BP; 716 A; 1137 C; 981 G; 647 T; 0 U; 1 Other;  
 XX  
 XX Query Match 2.2%; Score 27; DB 7; Length 3482;  
 XX Best Local Similarity 100.0%; Pred. No. 0.042;  
 XX Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 367 TACATCGGCTCATCGCCATGGCCATC 393  
 DB 832 TACATCGGCTCATCGCCATGGCCATC 858

RESULT 12  
 ABQ25892/c  
 ID ABQ25892 standard; DNA; 738 BP.  
 XX AC ABQ25892;  
 XX  
 XX 12-JUL-2002 (first entry)  
 XX  
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 12483.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200218632-A2.  
 XX  
 XX 07-MAR-2002.  
 XX  
 XX 01-SEP-2001; 2001WO-EP010074.  
 XX  
 XX 01-SEP-2000; 2000DE-01043826.  
 XX 05-SEP-2000; 2000DE-01044543.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K, Guestig D;  
 XX  
 XX WPI; 2002-371829/40.  
 XX  
 XX Determining the degree of cytosine methylation in genomic DNA, useful for  
 XX diagnosis and prognosis, comprises selective hybridization of amplicons  
 XX from chemically treated DNA.  
 XX  
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 XX This invention describes a novel method for determining the degree of  
 XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 XX genomic sample of DNA. The sample is treated chemically to convert  
 XX cytosine (C) but not methylated C, to uracil, then part of the genomic  
 XX DNA that contains the target C is amplified to form a labeled amplicon.  
 XX The amplicon is hybridised to two classes, each with at least one member,  
 XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
 XX degree of hybridisation to both classes is determined from the label on  
 XX the amplicon. From the ratio of labels hybridised to the two classes of  
 XX oligomers, the degree of methylation is calculated. The method is used:  
 XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
 XX and of a wide range of diseases, e.g. cancer, disorders of the central

CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
 CC particularly by detecting mutations or single nucleotide polymorphisms  
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
 CC investigating cell differentiation. The method allows the methylation  
 CC status of many C residues to be determined simultaneously. ABQ13410-  
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
 CC for determining the degree of cytosine methylation described in the  
 CC disclosure of the invention

CC Sequence 738 BP; 106 A; 63 C; 274 G; 295 T; 0 U; 0 Other;  
 XX Query Match 2.2%; Score 26; DB 6; Length 738;  
 XX Best Local Similarity 100.0%; Pred. No. 0.13;  
 XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 356 CCCCTACTCGTACATCGCGTCTATC 381  
 Db 373 CCCCTACTCGTACATCGCGTCTATC 348

RESULT 13  
 ABQ25893  
 ID ABQ25893 standard; DNA; 738 BP.  
 XX AC ABQ25893;  
 XX DT 12-JUL-2002 (first entry)  
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 12484.  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.  
 XX OS Homo sapiens.  
 XX FN WO200218632-A2.  
 XX PD 07-MAR-2002.  
 XX PF 01-SEP-2001; 2001WO-EP010074.  
 XX PR 01-SEP-2000; 2000DE-01043826.  
 XX PR 05-SEP-2000; 2000DE-01044543.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX WPI; 2002-371829/40.  
 XX Determining the degree of cytosine methylation in genomic DNA, useful for  
 XX diagnosis and prognosis, comprises selective hybridization of amplicons  
 XX from chemically treated DNA.  
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of  
 methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 genomic sample of DNA. The sample is treated chemically to convert  
 cytosine (C) but not methylated C, to uracil, then part of the genomic  
 DNA that contains the target C is amplified to form a labeled amplicon.  
 The amplicon is hybridised to two classes, each with at least one member,  
 of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
 degree of hybridisation to both classes is determined from the label on  
 the amplicon. From the ratio of labels hybridised to the two classes of  
 oligomers, the degree of methylation is calculated. The method is used:  
 (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
 and of a wide range of diseases, e.g. cancer, disorders of the central  
 nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
 particularly by detecting mutations or single nucleotide polymorphisms  
 (SNP's); and (ii) for differentiation of cell or tissue types and for

CC investigating cell differentiation. The method allows the methylation  
 CC status of many C residues to be determined simultaneously. ABQ13410-  
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
 CC for determining the degree of cytosine methylation described in the  
 CC disclosure of the invention

CC Sequence 738 BP; 295 A; 274 C; 63 G; 106 T; 0 U; 0 Other;  
 XX Query Match 2.2%; Score 26; DB 6; Length 738;  
 XX Best Local Similarity 100.0%; Pred. No. 0.13;  
 XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 356 CCCCTACTCGTACATCGCGTCTATC 381  
 Db 366 CCCCTACTCGTACATCGCGTCTATC 391

RESULT 14  
 ABL80120  
 ID ABL80120 standard; cDNA; 441 BP.  
 XX AC ABL80120;  
 XX DT 17-MAY-2002 (first entry)  
 XX DE Human ovarian cancer related cDNA clone SEQ ID NO:3098.  
 XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
 XX OS Homo sapiens.  
 XX PN WO200192581-A2.  
 XX PD 06-DEC-2001.  
 XX PF 29-MAY-2001; 2001WO-US017756.  
 XX PR 26-MAY-2000; 2000US-0207484P.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Algate PA, Harlocker SL, Jones R;  
 XX WPI; 2002-122075/16.  
 XX Composition for therapy and diagnosis of ovarian cancer comprising  
 XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
 XX polypeptide, antibody specific to polypeptide or T cell expressing  
 XX polypeptide.  
 XX Claim 1; SEQ ID NO 3098; 489pp; English.

The present invention describes a composition (I) comprising: carriers  
 and immunostimulants; and a polypeptide (II) of a ovarian tumour  
 polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)  
 from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,  
 (III) encoding (II) having a sequence (S2), a T cell population of (II),  
 or antigen presenting cells that express (II). (I) has cytostatic  
 activity. An oligonucleotide (IV) that hybridises to (S1) can be used for  
 detecting ovarian cancer in a patient's biological sample preferably  
 serum or ovarian tissue. The method comprises contacting a biological  
 sample from a patient with (IV), detecting the amount of polynucleotide  
 hybridising to (IV) and comparing the amount to a predetermined cutoff  
 value and thereby detecting ovarian cancer in the patient, where the  
 amount of polynucleotide hybridising to (IV) is detected preferably by  
 polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is  
 useful for stimulating and/or expanding T cells specific for an ovarian  
 tumour protein comprising contacting T cells with (III) or (II). (III) is  
 useful in design and preparation of ribozyme molecules for inhibiting  
 expression of the tumour polypeptides and proteins in tumour cells; and  
 to isolate a full length gene from a suitable library e.g., a tumour cDNA  
 library using well known techniques

SQ Sequence 441 BP; 95 A; 116 C; 96 G; 134 T; 0 U; 0 Other;  
Query Match 2.1%; Score 25; DB 6; Length 441;  
Best Local Similarity 100.0%; Pred. No. 0.39;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1181 CGTACCGGTGGAGAGCTCCTAGC 1205  
Db 13 CGTACCGGTGGAGAGCTCCTAGC 37

Best Local Similarity 100.0%; Pred. No. 0.37;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 358 CCCTACTCGTACATCGGCTCATCG 382  
Db 59 CCCTACTCGTACATCGGCTCATCG 83  
Search completed: March 31, 2004, 02:34:43  
Job time : 507 secs

RESULT 15  
ABQ34485  
ID ABQ34485 standard; DNA; 890 BP.  
XX AC ABQ34485;  
XX AC ABQ34485;  
DT 12-JUL-2002 (first entry)  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 21076.  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX Homo sapiens.  
OS Homo sapiens.  
XX WO200218632-A2.  
PN WO200218632-A2.  
XX 07-MAR-2002.  
PD 07-MAR-2002.  
XX 01-SEP-2001; 2001WO-EP010074.  
XX 01-SEP-2000; 2000DE-01043826.  
PR 05-SEP-2000; 2000DE-0104543.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
PI WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ3410-  
CC ABQ34121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX Sequence 890 BP; 311 A; 354 C; 106 G; 119 T; 0 U; 0 Other;

Query Match 2.1%; Score 25; DB 6; Length 890;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 02:18:33 ; Search time 114 Seconds  
(without alignments)  
5885.400 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209  
Sequence: 1 atgaagtggagggttcgt.....tggagagctcctagcttga 1209

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	6.6	320	4	US-09-833-381-1266
2	25	2.1	1965	4	US-09-220-132-21
3	25	2.1	4695	2	US-08-231-193A-57
4	25	2.1	4695	2	US-08-486-273A-57
5	25	2.1	4695	3	US-08-940-086A-57
6	25	2.1	4695	4	US-08-940-035A-57
7	25	2.1	4695	4	US-08-935-105A-57
8	25	2.1	4695	4	US-09-648-797-57
9	25	2.1	4695	4	US-09-386-123-57
10	24	2.0	159	3	US-09-437-457-11
11	24	2.0	429	4	US-09-641-638-324
12	24	2.0	1860	2	US-08-331-644-3
13	24	2.0	1860	5	PCT-US93-04102-3
14	24	2.0	3387	1	US-08-064-271-11
15	24	2.0	3387	2	US-08-627-254C-29
16	24	2.0	3387	4	US-09-023-655-1050
17	22	1.8	3387	3	US-08-930-589A-19
18	22	1.8	3387	4	US-09-599-781-19
19	20	1.7	480	4	US-09-674-608A-4
20	20	1.7	1454	4	US-09-372-422A-19
21	20	1.7	2561	4	US-09-616-289-48
22	20	1.7	4403765	3	US-09-103-840A-2
23	20	1.7	4411529	3	US-09-103-840A-1
24	19	1.6	576	4	US-09-252-991A-13112
25	19	1.6	633	4	US-09-252-991A-12687
26	19	1.6	720	4	US-09-252-991A-2351
27	19	1.6	720	4	US-09-252-991A-2640

C	28	19	1.6	1101	4	US-09-489-039A-2638	Sequence 2638, Ap
	29	19	1.6	1192	3	US-09-318-448-18	Sequence 18, Appl
C	30	19	1.6	1248	4	US-09-489-039A-6144	Sequence 6144, Ap
	31	19	1.6	1278	4	US-09-252-991A-2483	Sequence 2483, Ap
	32	19	1.6	1563	4	US-09-252-991A-2566	Sequence 2566, Ap
	33	19	1.6	2124	4	US-09-252-991A-12830	Sequence 12830, A
	34	19	1.6	2634	3	US-08-941-936-1	Sequence 1, Appli
C	35	19	1.6	3623	1	US-08-306-691B-35	Sequence 1, Appli
	36	19	1.6	43280	2	US-08-804-227C-1	Sequence 1, Appli
C	37	18	1.5	18	2	US-08-627-254C-12	Sequence 12, Appl
	38	18	1.5	248	4	US-09-397-787-24	Sequence 24, Appl
	39	18	1.5	420	4	US-09-252-991A-7967	Sequence 7967, Ap
	40	18	1.5	504	4	US-09-252-991A-9584	Sequence 9584, Ap
	41	18	1.5	957	4	US-09-252-991A-51	Sequence 51, Appl
	42	18	1.5	1077	4	US-09-522-714-17	Sequence 17, Appl
C	43	18	1.5	1143	4	US-09-252-991A-5748	Sequence 5748, Ap
	44	18	1.5	1188	4	US-09-252-991A-5131	Sequence 5131, Ap
C	45	18	1.5	1239	4	US-09-252-991A-9666	Sequence 9666, Ap

ALIGNMENTS

RESULT 1  
US-09-833-381-1266  
; Sequence 1266, Application US/09833381  
; Patent No. 6672186  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs  
; FILE REFERENCE: 5800-119  
; CURRENT APPLICATION NUMBER: US/09/833,381  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 09/516,448  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 2050  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1266  
; LENGTH: 320  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(320)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-833-381-1266

Query Match	6.6%	Score 80;	DB 4;	Length 320;
Best Local Similarity	100.0%	Pred. No. 1.4e-24;		
Matches	80;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	567	GGCACTACTGATGCTCAACCCCAACGCGAGTACACCTTCGCGACGGGCTCTTCG	826	
Db	1	GGCACTACTGATGCTCAACCCCAACGCGAGTACACCTTCGCGACGGGCTCTTCG	60	
Qy	627	CCGCGCGCGCAAGCGCTCA	646	
Db	61	CCGCGCGCGCAAGCGCTCA	80	

RESULT 2  
US-09-220-132-21  
; Sequence 21, Application US/09220132  
; Patent No. 6506607  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER.  
; FILE REFERENCE: 07334-074001  
; CURRENT APPLICATION NUMBER: US/09/220,132  
; CURRENT FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: US 60/079,303  
; PRIOR FILING DATE: 1998-03-25

;; PRIOR APPLICATION NUMBER: US 60/068,821  
;; PRIOR FILING DATE: 1997-12-24  
;; NUMBER OF SEQ ID NOS: 191  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 21  
;; LENGTH: 1965  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-220-132-21

Query Match 2.1%; Score 25; DB 4; Length 1965;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 CCCTACTGCTACATCGCGCTCATCG 382  
DB 196 CCCTACTGCTACATCGCGCTCATCG 220

RESULT 3  
US-08-231-193A-57  
; Sequence 57, Application US/08231193A  
; Patent No. 5849895  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231.193A  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION NUMBER: US 08/052,459  
; FILING DATE: 20-APR-1993  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4695 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 485..4495  
US-08-231-193A-57

Query Match 2.1%; Score 25; DB 2; Length 4695;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGGGG 309  
DB 3700 GGGGCCAGCGCGCGCGGGG 3724

RESULT 4  
US-08-486-273A-57  
; Sequence 57, Application US/08486273A  
; Patent No. 5985586  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA  
; TITLE OF INVENTION: ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486.273A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/231.193  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4695 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 485..4495  
US-08-486-273A-57

Query Match 2.1%; Score 25; DB 2; Length 4695;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGGGG 309  
DB 3700 GGGGCCAGCGCGCGCGGGG 3724

RESULT 5  
US-08-940-086A-57  
; Sequence 57, Application US/08940086A  
; Patent No. 611091  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.

```
/ APPLICANT: Liaw, Chen W.
/ APPLICANT: Lu, Chin-Chun
/ TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
/ TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
/ NUMBER OF SEQUENCES: 63
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Heller Ehrman White & McAuliffe
/ STREET: 4250 Executive Square, 7th Floor
/ CITY: La Jolla
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 92037
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/940,086A
/ FILING DATE: 29-SEPT-97
/ CLASSIFICATION: 536
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/231,193
/ FILING DATE: 20-APR-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/052,449
/ FILING DATE: 20-APR-1993
/ NAME: Seidman, Stephanie
/ REGISTRATION NUMBER: 33,779
/ REFERENCE/DOCKET NUMBER: 24735-9383C
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 450-8400
/ TELEFAX: (619) 450-8499
/ INFORMATION FOR SEQ ID NO: 57:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4695 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: both
/ TOPOLOGY: both
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 485..4495
/ US-08-940-086A-57

Query Match 2.1%; Score 25; DB 3; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGCGG 309
Db 3700 GGGGCCAGCGCGCGCGCGGG 3724

RESULT 6
US-08-940-035A-57
; Sequence 57, Application US/08940035A
; Patent No. 6316611
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.

APPLICANT: Liaw, Chen W.
APPLICANT: Lu, Chin-Chun
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehrman White & McAuliffe
STREET: 4250 Executive Square, 7th Floor
CITY: La Jolla
STATE: CA
COUNTRY: U.S.A.
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/940,086A
FILING DATE: 29-SEPT-97
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/231,193
FILING DATE: 20-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/052,449
FILING DATE: 20-APR-1993
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24735-9383C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 450-8400
TELEFAX: (619) 450-8499
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 4695 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 485..4495
US-08-940-086A-57

Query Match 2.1%; Score 25; DB 3; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGCGG 309
Db 3700 GGGGCCAGCGCGCGCGGG 3724

RESULT 7
US-08-935-105A-57
; Sequence 57, Application US/08935105A
; Patent No. 6376660
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/935,105A
; FILING DATE: 29-SEPT-97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; PRIOR APPLICATION DATA:
```

```
/ ZIP: 92037
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/940,035A
/ FILING DATE: 29-SEPT-97
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/231,193
/ FILING DATE: 20-APR-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/052,449
/ FILING DATE: 20-APR-1993
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seidman, Stephanie
/ REGISTRATION NUMBER: 33,779
/ REFERENCE/DOCKET NUMBER: 6362-9383E
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 619-238-0999
/ TELEFAX: 619-238-0062
/ INFORMATION FOR SEQ ID NO: 57:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4695 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: both
/ TOPOLOGY: both
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 485..4495
/ US-08-940-035A-57

Query Match 2.1%; Score 25; DB 4; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGGG 309
Db 3700 GGGGCCAGCGCGCGCGGG 3724

RESULT 7
US-08-935-105A-57
; Sequence 57, Application US/08935105A
; Patent No. 6376660
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/935,105A
; FILING DATE: 29-SEPT-97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; PRIOR APPLICATION DATA:
```

```
/ APPLICATION NUMBER: US 08/052,449
/ FILING DATE: 20-APR-1993
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seidman, Stephanie
/ REGISTRATION NUMBER: 33,779
/ REFERENCE/DOCKET NUMBER: 6362-9383D
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 619-238-0999
/ TELEFAX: 619-238-0062
/ INFORMATION FOR SEQ ID NO: 57:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4695 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: both
/ TOPOLOGY: both
/ MOLECULE TYPE: CDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 485..4495
/ SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-08-648-797-57

Query Match          2.1%; Score 25; DB 4; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGGCGCGCGCGCGCGGG 309
Db 3700 GGGGCCAGGCGCGCGCGCGCGGG 3724

RESULT 9
US-09-386-123-57
/ Sequence 57, Application US/09386123
/ Patent No. 6521413
/ GENERAL INFORMATION:
/ APPLICANT: Daggett, Lorrie P.
/ APPLICANT: Ellis, Steven B.
/ APPLICANT: Liaw, Chen W.
/ APPLICANT: Lu, Chin-Chun
/ TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
/ NUMBER OF SEQUENCES: 63
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Heller Ehrman White & McAuliffe
/ STREET: 4250 Executive Square, 7th Floor
/ CITY: La Jolla
/ STATE: CA
/ COUNTRY: U.S.A.
/ ZIP: 92037
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/386,123
/ FILING DATE:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/486,273
/ FILING DATE: 06-JUNE-95
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/231,193
/ FILING DATE: 20-APR-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/052,449
/ FILING DATE: 20-APR-1993
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seidman, Stephanie
/ REGISTRATION NUMBER: 33,779
/ REFERENCE/DOCKET NUMBER: 6362-9383F
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 858-450-8403
/ TELEFAX: 858-587-5360
/ INFORMATION FOR SEQ ID NO: 57:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4695 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: both
/ TOPOLOGY: both
/ MOLECULE TYPE: CDNA
/ FEATURE:
/
```

```
; NAME/KEY: CDS
; LOCATION: 485..4495
US-09-386-123-57

Query Match      2.1%; Score 25; DB 4; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGGGCGGGCGGGG 309
      |||||
Db 3700 GGGGCCAGCGCGGGCGGGCGGGG 3724

RESULT 10
US-09-437-457-11
; Sequence 11, Application US/09437457
; Patent No. 6273893
; GENERAL INFORMATION:
; APPLICANT: Giordano, Anthony
; APPLICANT: Xavier, Ashish
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES AND METHODS FOR
; TITLE OF INVENTION: IDENTIFYING COMPOUNDS THAT AFFECT RNA/RNA BINDING PROTEIN
; TITLE OF INVENTION: INTERACTIONS AND MRNA FUNCTIONALITY
; FILE REFERENCE: 50093/014001
; CURRENT APPLICATION NUMBER: US/09/437,457
; CURRENT FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-437-457-11

Query Match      2.0%; Score 24; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 GCGCGCGCGCTCGCGCGCGCTGC 913
      |||||
Db 73 GCGCGCGCGCTCGCGCGCGCTGC 96

RESULT 11
US-09-641-638-324
; Sequence 324, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 324
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
```

```
; LOCATION: 386
; OTHER INFORMATION: 10-52-386 : polymorphic base C or T
; NAME/KEY: misc_binding
; LOCATION: 366..385
; OTHER INFORMATION: 10-52-386.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 387..406
; OTHER INFORMATION: 10-52-386.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 1..18
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 412..429
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 374..398
; OTHER INFORMATION: 10-52-386 potential probe
US-09-641-638-324

Query Match      2.0%; Score 24; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 GCGCGCGCGCTCGCGCGCGCTGC 913
      |||||
Db 160 GCGCGCGCGCTCGCGCGCGCTGC 183

RESULT 12
US-08-331-644-3
; Sequence 3, Application US/08331644
; Patent No. 5976872
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,644
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,292
; FILING DATE: 13-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1860 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-331-644-3
```

Query Match 2.0%; Score 24; DB 2; Length 1860;  
 Best Local Similarity 100.0%; Pred. No. 0.34;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 CCCTACTCGTACATCGCGCTCATC 381  
 DB 759 CCCTACTCGTACATCGCGCTCATC 782

RESULT 13

PCT-US93-04102-3  
 ; Sequence 3, Application PC/TUS9304102  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tao, Wufan  
 ; APPLICANT: Lai, Eseng  
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White  
 ; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/04102  
 ; FILING DATE: 19930430  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/882,292  
 ; FILING DATE: 13-MAY-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 41472A-PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-977-9550  
 ; TELEFAX: 212-664-0525  
 ; TELEX: 422523 COOP UI  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1860 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cdna  
 ; PCT-US93-04102-3

Query Match 2.0%; Score 24; DB 5; Length 1860;  
 Best Local Similarity 100.0%; Pred. No. 0.34;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 CCCTACTCGTACATCGCGCTCATC 381  
 DB 759 CCCTACTCGTACATCGCGCTCATC 782

RESULT 14

US-08-084-271-11  
 ; Sequence 11, Application US/08064271  
 ; Patent No. 5543297  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kennedy, Brian P.  
 ; APPLICANT: Cromlish, Wanda A.  
 ; APPLICANT: Mancini, Joseph A.  
 ; APPLICANT: O'Neill, Gary  
 ; APPLICANT: Vickers, Phillip J.

APPLICANT: Wong, Elizabeth  
 TITLE OF INVENTION: HUMAN CYCLOOXYGENASE-2 CDNA AND  
 TITLE OF INVENTION: ASSAY FOR EVALUATING CYCLOOXYGENASE ACTIVITY  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Merck & Co., Inc.  
 STREET: 126 Lincoln Avenue  
 CITY: Rahway  
 STATE: NJ  
 COUNTRY: USA  
 ZIP: 07065

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 in, 1.4kb  
 COMPUTER: Apple Macintosh  
 OPERATING SYSTEM: System 7  
 SOFTWARE: Microsoft Word 5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/064,271  
 FILING DATE: 19930506  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Panzer, Curtis C.  
 REGISTRATION NUMBER: 33,752  
 REFERENCE/DOCKET NUMBER: 189061A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (908)594-3199  
 TELEFAX: (908)594-4720  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3387 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-064-271-11

Query Match 2.0%; Score 24; DB 1; Length 3387;  
 Best Local Similarity 100.0%; Pred. No. 0.32;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 GGGCGCGCGCTGCGCGCGCTGC 913  
 DB 73 GGGCGCGCGCTGCGCGCGCTGC 96

RESULT 15

US-08-627-254C-29  
 ; Sequence 29, Application US/08627254C  
 ; Patent No. 5859229  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kniss, Douglas A.  
 ; TITLE OF INVENTION: Eicosanoid Formation  
 ; NUMBER OF SEQUENCES: 29  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Calfee, Halter & Griwold LLP  
 ; STREET: 800 Superior Avenue  
 ; CITY: Cleveland  
 ; STATE: Ohio  
 ; COUNTRY: USA  
 ; ZIP: 44114  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/627,254C  
 ; FILING DATE:  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Goldrick, Mary E  
 ; REGISTRATION NUMBER: 34,829  
 ; REFERENCE/DOCKET NUMBER: 18525/00107

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3387 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
US-08-627-254C-29

Query Match      2.0%; Score 24; DB 2; Length 3387;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      890 GCGCCGCGCCCTGCGCCGCGCTGC 913
Db      73 GCGCCGCGCCCTGCGCCGCGCTGC 96

Search completed: March 31, 2004, 04:45:16
JOB time : 132 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 03:52:59 ; Search time 2471 Seconds  
(without alignments)  
1821.578 Million cell up

Title: US-10-087-080-31

Perfect score: 1209  
Sequence: 1 atcaagattgagagttgttcgt.....tggagacgctctctagcttga 1209

Scoring table: OLIGO · NUC

Gapop 60.0 , Gapext 60.0

Searched: 2458946 seqs, 1861504846 residues

Word size :

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:\*

1:	/cgn2_6/pdata1/1/pubpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/pdata1/1/pubpna/US07_PUBCOMB.seq.*
3:	/cgn2_6/pdata1/1/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/pdata1/1/pubpna/US06_NEW_PUB.seq.*
5:	/cgn2_6/pdata1/1/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/pdata1/1/pubpna/US07_NEW_PUB.seq.*
7:	/cgn2_6/pdata1/1/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/pdata1/1/pubpna/US08_NEW_PUB.seq.*
9:	/cgn2_6/pdata1/1/pubpna/US09A_PUBCOMB.seq.*
10:	/cgn2_6/pdata1/1/pubpna/US09B_PUBCOMB.seq.*
11:	/cgn2_6/pdata1/1/pubpna/US09C_PUBCOMB.seq.*
12:	/cgn2_6/pdata1/1/pubpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/pdata1/1/pubpna/US10A_PUBCOMB.seq.*
14:	/cgn2_6/pdata1/1/pubpna/US10B_PUBCOMB.seq.*
15:	/cgn2_6/pdata1/1/pubpna/US10C_PUBCOMB.seq.*
16:	/cgn2_6/pdata1/1/pubpna/US10_NEW_PUB.seq.*
17:	/cgn2_6/pdata1/1/pubpna/US60_NEW_PUB.seq.*
18:	/cgn2_6/pdata1/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1209	100.0	1209	12	US-10-229-345-9	Sequence 9, Appli
2	1209	100.0	1209	12	US-10-274-177-9	Sequence 9, Appli
3	1209	100.0	1209	15	US-10-087-080-31	Sequence 31, Appl
C 4	264	21.8	585	15	US-10-037-632-231353	Sequence 231353,
5	80	6.6	320	9	US-09-833-381-1266	Sequence 1266, Ap
6	30	2.5	312	14	US-10-029-386-23708	Sequence 23708, A
7	30	2.5	425	15	US-10-037-632-138699	Sequence 138699,
8	30	2.5	425	15	US-10-037-632-138700	Sequence 138700,
9	30	2.5	425	15	US-10-027-632-138701	Sequence 138701,
10	30	2.5	425	15	US-10-027-632-138703	Sequence 138703,
11	30	2.5	599	14	US-10-039-386-10008	Sequence 10008, A
12	25	2.1	441	9	US-09-867-701-3098	Sequence 3098, Ap
13	25	2.1	1335	14	US-10-029-386-20506	Sequence 20506, A
C 14	25	2.1	2187	10	US-09-960-706-947	Sequence 947, App
15	25	2.1	4695	9	US-09-945-901-57	Sequence 57, Appl

16	25	2.1	4695	13	US-10-007-747-57	Sequence 57, Appl
17	25	2.1	4695	14	US-10-038-937-57	Sequence 57, Appl
18	24	2.0	159	9	US-09-729-447-11	Sequence 11, Appl
19	24	2.0	159	9	US-09-437-958-11	Sequence 11, Appl
20	24	2.0	310	15	US-10-393-403-26	Sequence 26, Appl
21	24	2.0	4056	14	US-10-107-097-324	Sequence 324, App
22	24	2.0	10256	15	US-10-101-510-419	Sequence 419, App
23	24	2.0	2271	14	US-10-071-510-133	Sequence 133, App
24	24	2.0	2487	14	US-10-205-823-135	Sequence 135, App
25	24	2.0	2563	13	US-10-044-090-477	Sequence 477, App
26	24	2.0	3387	15	US-10-191-597-87	Sequence 87, Appl
27	24	2.0	4496	13	US-10-044-090-478	Sequence 478, App
28	24	2.0	4496	13	US-10-071-766-46	Sequence 46, Appl
29	24	2.0	4750	14	US-10-247-671-136	Sequence 126, App
30	24	2.0	11064	10	US-09-949-293-25	Sequence 25, Appl
31	23	1.9	977	14	US-10-029-386-24193	Sequence 24193, A
32	23	1.9	996	14	US-10-156-761-2841	Sequence 2841, Ap
33	23	1.9	2314	15	US-10-108-761-1673	Sequence 1673, Ap
34	23	1.9	9025608	14	US-10-152-761-1	Sequence 1, Appl
35	22	1.8	597	14	US-10-029-386-20713	Sequence 20713, A
36	22	1.8	958	12	US-10-425-114-649	Sequence 649, App
37	22	1.8	3289	9	US-09-963-285-8	Sequence 8, Appl
38	22	1.8	3387	14	US-10-027-961A-19	Sequence 19, Appl
39	22	1.8	4024	15	US-10-295-027-1039	Sequence 1039, Ap
40	22	1.8	4045	14	US-10-161-572-20	Sequence 20, Appl
41	22	1.8	4465	10	US-09-953-067A-2	Sequence 2, Appl
42	22	1.8	4465	12	US-10-211-463-132	Sequence 132, App
43	22	1.8	4465	14	US-10-021-660-42	Sequence 42, Appl
44	22	1.8	4465	15	US-10-395-027-31	Sequence 31, Appl
45	22	1.8	4465	15	US-10-273-801-11	Sequence 11, Appl

## ALIGNMENTS

```

RESULT 1
US-10-229-345-9
; Sequence 9, Application US/1029345
; Publication No. US20040038220A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CRWU-P01-003
; CURRENT APPLICATION NUMBER: US/10/229,345
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1161)..(1161)
; OTHER INFORMATION: n=a, c, g, or t
US-10-229-345-9

```

	Query Match	100.0%;	Score 1209;	DB 12;	Length 1209;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 1209;	Conservative	0;	Mismatches	0;
				Indels	Gaps
Qy	1	ATGAAGTTGGAGTGTTCTGTCCTTCGGCGGCCACCGGGACAGAGGCGAGTGACTG	60		
Db	1	ATGAAGTTGGAGTGTTCTGTCCTTCGGCGGCCACCGGGACAGAGGCGAGTGACTG	60		
Qy	61	GAGCGGCGCGGCGGACGACGCGCGCTCCCGCTGTCTCGCGCGGAGACGACTCCCTG	120		
Db	61	GAGCGGCGCGGCGGACGACGCGCGCTCCCGCTGTCTCGCGCGGAGACGACTCCCTG	120		
Qy	121	GGTTCAGATGGGACTCCGCGGCAAGCCGTCGCGCGGCGCGCGCCAGAGATACGCAG	180		
Db	121	GGTTCAGATGGGACTCCGCGGCAAGCCGTCGCGCGGCGCGCGCCAGAGATACGCAG	180		

```
181 GCGACGCGGCAACAGAGTGGGAGGCGGCGCGGCGGAGGAGCGGATCCGGCAGCA 240
181 GCGACGCGGCAACAGAGTGGGAGGCGGCGGCGGAGGAGGCGGATCCGGCAGCA 240
241 GCTGCTGACGCGTGGTGGGAGGCGCGGAGGCGGCGGCGGCGGCGGCGGCGG 300
241 GCTGCTGACGCGTGGTGGGAGGCGCGGAGGCGGCGGCGGCGGCGGCGGCGG 300
301 GCGCGCGGAGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 360
301 GCGCGCGGAGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 360
361 TACTGCTACATCGCGTTCATCGCATGGCCATCGGCACTCGGCGGCGGCGGCGG 420
361 TACTGCTACATCGCGTTCATCGCATGGCCATCGGCACTCGGCGGCGGCGGCGG 420
421 CTGCGGAGATCAACGAGTACCTCATGGCAAGTTCCTCTTTTCGCGGCGGAGTACG 480
421 CTGCGGAGATCAACGAGTACCTCATGGCAAGTTCCTCTTTTCGCGGCGGAGTACG 480
481 GGCTGGCGCAACTCCGTGGCGCCCAACCTTTTCGCTCAACGACTGCTTCGAGTGTG 540
481 GGCTGGCGCAACTCCGTGGCGCCCAACCTTTTCGCTCAACGACTGCTTCGAGTGTG 540
541 CGGACCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
541 CGGACCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
601 TACACCTTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
601 TACACCTTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
661 GTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720
661 GTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720
721 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 780
721 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 780
781 AGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 840
781 AGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 840
841 CGCAGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
841 CGCAGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
901 TGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 960
901 TGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 960
961 CCGCTCTGCGGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1020
961 CCGCTCTGCGGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1020
1021 CCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
1021 CCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
1081 CTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1140
1081 CTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1140
1141 CTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
1141 CTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
1201 CTAGCTTGA 1209
1201 CTAGCTTGA 1209
```

```
RESULT 2
US-10-274-177-9
; Sequence 9, Application US/10274177
; Publication No. US20040038225A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/274,177
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US/10/229,345
; PRIOR FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1161)..(1161)
; OTHER INFORMATION: n=a, c, g, or t
US-10-274-177-9

Query Match      100.0%; Score 1209; DB 12; Length 1209;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTTGGAGGTTCGTCCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
DB 1 ATGAAGTTGGAGGTTCGTCCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
QY 61 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 120
DB 61 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 120
QY 121 GGTCTCAGATGGGAGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 180
DB 121 GGTCTCAGATGGGAGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 180
QY 181 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 240
DB 181 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 240
QY 241 GCTGCTGAGCGGTTCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
DB 241 GCTGCTGAGCGGTTCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
QY 301 GCGCGGCGGAGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 360
DB 301 GCGCGGCGGAGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 360
QY 361 TACTGCTACATCGGCTCATCGCCATGGCCATCGGCACTCGGCGGCGGCGGCGGCG 420
DB 361 TACTGCTACATCGGCTCATCGCCATGGCCATCGGCACTCGGCGGCGGCGGCGGCG 420
QY 421 CTGCGGAGATCAACGAGTACCTCATGGCAAGTTCCTCTTTTCGCGGCGGAGTACG 480
DB 421 CTGCGGAGATCAACGAGTACCTCATGGCAAGTTCCTCTTTTCGCGGCGGAGTACG 480
QY 481 GGTCTGCGCAACTCCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 540
DB 481 GGTCTGCGCAACTCCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 540
QY 541 GCGGACCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 541 GCGGACCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
QY 601 TACACCTTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
DB 601 TACACCTTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
QY 661 GTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720
```

Db 661 GTCCCGCGCCGCGGCTGCGGCGCGAGAGCGCCCGGCGCTCCCGCGCGCCCGCGCC 720  
Qy 721 GCGCCCGCCCGCGGCTCGCCCGCATCGCTCGCCCGCGCGCGCGCGAGAGCGCGCC 780  
Db 721 GCGCCCGCGCCCGCGGCTCGCCCGCATCGCTCGCCCGCGCGCGCGAGAGCGCGCC 780  
Qy 781 AGCCCGCGGCGAGTCTCCAGCTCTTGGCCATCGACAGCATCTGCGCAGCCCTTC 840  
Db 781 AGCCCGCGGCGAGTCTCCAGCTCTTGGCCATCGACAGCATCTGCGCAGCCCTTC 840  
Qy 841 CGCAGCGCTCGCTCAGGAGACAGCGCCCGCGGACGACGCTTCAGTGGGCGCGCGCC 900  
Db 841 CGCAGCGCTCGCTCAGGAGACAGCGCCCGCGGACGACGCTTCAGTGGGCGCGCGCC 900  
Qy 901 TGCCCGCGCTGCGCGGCTTCCCGGCGCTCTCCCGCGCGCGCGCTTCAGGCGCTGCTG 960  
Db 901 TGCCCGCGCTGCGCGGCTTCCCGGCGCTCTCCCGCGCGCGCGCTTCAGGCGCTGCTG 960  
Qy 961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCTGCGCGCGCGAGGCTG 1020  
Db 961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCTGCGCGCGCGAGGCTG 1020  
Qy 1021 CCACCGACCGCGCGCCCTCTCTGTTGCACTCTCCCGCGCGCGCGCGCGCGCGCG 1080  
Db 1021 CCACCGACCGCGCGCCCTCTCTGTTGCACTCTCCCGCGCGCGCGCGCGCGCGCG 1080  
Qy 1081 CTCCGAGCG 1140  
Db 1081 CTCCGAGCG 1140  
Qy 1141 CTCCGAGCG 1200  
Db 1141 CTCCGAGCG 1200  
Qy 1201 CTAGCTTGA 1209  
Db 1201 CTAGCTTGA 1209

RESULT 3  
US-10-087-080-31  
; Sequence 31, Application US/10097080  
; Publication No. US20030235820A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Markowitz, Sanford David  
; APPLICANT: Eos Biotechnology, Inc.  
; APPLICANT: Case Western Reserve University  
; TITLE OF INVENTION: No. US20030235820A1 Methods of Diagnosis of Metastatic Colorectal Cancer  
; TITLE OF INVENTION: Cancer, Compositions and Methods of Screening for  
; TITLE OF INVENTION: Modulators of Metastatic Colorectal Cancer  
; FILE REFERENCE: 018501-000840US  
; CURRENT APPLICATION NUMBER: US/10/087,080  
; PRIOR FILING DATE: 2002-10-25  
; PRIOR FILING DATE: 2001-02-27, 206  
; PRIOR FILING DATE: 2001-02-27  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR FILING DATE: 2001-04-17  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 31  
; LENGTH: 1209  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: winged helix/forkhead transcription factor (HFH1)  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (1161)  
; OTHER INFORMATION: n = g, a, c or t

US-10-087-080-31  
Query Match 100.0%; Score 1209; DB 15; Length 1209;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGAAGTTGGAGGTGTTCTGCTCCCTCGCGCGCGCCACAGGGGACAAAGCAGGCGAGTACCTG 60  
Db 1 ATGAAGTTGGAGGTGTTCTGCTCCCTCGCGCGCGCCACAGGGGACAAAGCAGGCGAGTACCTG 60  
Qy 61 GAGGCGCGCGCGCGCGCGCGCTCCCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 61 GAGGCGCGCGCGCGCGCGCGCTCCCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Qy 121 GGTCTCAGATGGGACTGCGCGCGCGCAAGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
Db 121 GGTCTCAGATGGGACTGCGCGCGCGCAAGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
Qy 181 GCGCAGCGCGAACAAGTGTGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 181 GCGCAGCGCGAACAAGTGTGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Qy 241 GCTGTGCGAGCGGTGCG 300  
Db 241 GCTGTGCGAGCGGTGCG 300  
Qy 301 GCGCGCGGAGCGCGAGGTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360  
Db 301 GCGCGCGGAGCGCGAGGTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360  
Qy 361 TACTGTACATCGGCTCATCGCCATCGCGCATCGCGCATCGCGCGCGCGCGCGCGCGCGCGCG 420  
Db 361 TACTGTACATCGGCTCATCGCCATCGCGCATCGCGCATCGCGCGCGCGCGCGCGCGCGCGCG 420  
Qy 421 CTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTTTCGCGCGCGCGCGCGCGCG 480  
Db 421 CTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTTTCGCGCGCGCGCGCGCGCG 480  
Qy 481 GGTGCGCGCAACTCGGTGCGCGCAACTCTTTCGCTCAACGACTGCTTCTGTTCAAGTGTGCTG 540  
Db 481 GGTGCGCGCAACTCGGTGCGCGCAACTCTTTCGCTCAACGACTGCTTCTGTTCAAGTGTGCTG 540  
Qy 541 GCGGACCCCTCGCGGCGCTCGGCG 600  
Db 541 GCGGACCCCTCGCGGCGCTCGGCG 600  
Qy 601 TACACCTTTCG 660  
Db 601 TACACCTTTCG 660  
Qy 661 GTCCCG 720  
Db 661 GTCCCG 720  
Qy 721 GCG 780  
Db 721 GCG 780  
Qy 781 AGCCCG 840  
Db 781 AGCCCG 840  
Qy 841 CGCAGCGCTCGCTCAGGAGACAGCGCCCGCGGAGCGCTTCAGTGGGCGCGCGCGCGCGCGCG 900  
Db 841 CGCAGCGCTCGCTCAGGAGACAGCGCCCGCGGAGCGCTTCAGTGGGCGCGCGCGCGCGCGCG 900  
Qy 901 TGCCCGCGCTGCGCGGCTTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
Db 901 TGCCCGCGCTGCGCGGCTTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
Qy 961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020  
Db 961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020



OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5  
OTHER INFORMATION: SWISSPROT HIT: Q60688, EVALUE 1.00e-58  
OTHER INFORMATION: EST HUMAN HIT: AM236501.1, EVALUE 3.00e-76  
OTHER INFORMATION: NT HIT: U13223.1, EVALUE 0.00e+00  
US-10-029-386-23708

Query Match 2.5%; Score 30; DB 14; Length 312;  
Best Local Similarity 100.0%; Pred. No. 0.00037;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381  
Db 11 AAGCCCCCTACTCGTACATCGCGCTCATC 40

## RESULT 7

US-10-027-632-138699  
Sequence 138699, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 138699

LENGTH: 425

TYPE: DNA

ORGANISM: Human

US-10-027-632-138699

Query Match 2.5%; Score 30; DB 15; Length 425;  
Best Local Similarity 100.0%; Pred. No. 0.00035;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381  
Db 91 AAGCCCCCTACTCGTACATCGCGCTCATC 120

## RESULT 8

US-10-027-632-138700  
Sequence 138700, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 138700  
LENGTH: 425  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-138700

Query Match 2.5%; Score 30; DB 15; Length 425;  
Best Local Similarity 100.0%; Pred. No. 0.00035;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381  
Db 91 AAGCCCCCTACTCGTACATCGCGCTCATC 120

## RESULT 9

US-10-027-632-138701  
Sequence 138701, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 138701

LENGTH: 425

TYPE: DNA

ORGANISM: Human

US-10-027-632-138701

Query Match 2.5%; Score 30; DB 15; Length 425;  
Best Local Similarity 100.0%; Pred. No. 0.00035;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381  
Db 91 AAGCCCCCTACTCGTACATCGCGCTCATC 120

## RESULT 10

US-10-027-632-138703  
Sequence 138703, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138703
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138703

Query Match          2.5%; Score 30; DB 15; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381
DB 91 AAGCCCCCTACTCGTACATCGCGCTCATC 120

RESULT 11
US-10-029-386-10008
; Sequence 10008, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10008
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 9.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
; OTHER INFORMATION: NT HIT: U13223.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q60688, EVALUE 2.00e-63
; OTHER INFORMATION: EST_HUMAN HIT: AW236501.1, EVALUE 5.00e-76
US-10-029-386-10008

Query Match          2.5%; Score 30; DB 14; Length 599;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381
```

```
DB 248 AAGCCCCCTACTCGTACATCGCGCTCATC 277

RESULT 12
US-09-867-701-3098
; Sequence 3098, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3098
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-3098

Query Match          2.1%; Score 25; DB 9; Length 441;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1181 CGTACCGGTGGAGACGCTCTAGC 1205
DB 13 CGTACCGGTGGAGACGCTCTAGC 37

RESULT 13
US-10-029-386-20506/c
; Sequence 20506, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20506
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008403.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
; OTHER INFORMATION: NT HIT: g113650682, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q03391, EVALUE 1.00e-61
; OTHER INFORMATION: EST_HUMAN HIT: B1912691.1, EVALUE 1.20e-01
US-10-029-386-20506

Query Match          2.1%; Score 25; DB 14; Length 1335;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGGG 309
DB 796 GGGGCCAGCGCGCGCGGG 772
```

```
RESULT 14
US-09-960-706-947
; Sequence 947, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; TITLE OF INVENTION: Gene Expression Profiles
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 947
; LENGTH: 2187
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 UI3220
US-09-960-706-947

Query Match      2.1%; Score 25; DB 10; Length 2187;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 CCTACTCGTACATCGCGTCTATCG 382
    |||||
Db 418 CCTACTCGTACATCGCGTCTATCG 442

RESULT 15
US-09-945-901-57
; Sequence 57, Application US/09945901
; Patent No. US20020161215A1
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; Ellis, Steven B.
; Liaw, Chen W.
; Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/945,901
; FILING DATE: 24-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/940,035
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383E
; TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0082
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4695 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 485..4495
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-09-945-901-57

Query Match      2.1%; Score 25; DB 9; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.049;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGCGGG 309
    |||||
Db 3700 GGGGCCAGCGCGCGCGCGCGGG 3724

Search completed: March 31, 2004, 06:38:09
Job time : 2479 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 02:16:13 ; Search time 2983 Seconds  
(without alignments)  
12103.043 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggaggttggtt.....tggagacgtctactgcttga 1209

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estlin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_esthum:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vit:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vil:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	376	31.1	904	13	BQ922461 AGENCOURT
2	146	12.1	982	13	BQ430047 AGENCOURT
3	131	10.8	246	29	CE688385 tigr-gss-
4	86	7.1	406	12	BMS10453 iJ41n03.y

5	77	6.4	398	9	AI586081
6	42	3.5	559	9	AI189632
7	38	3.1	232	12	BI402320
8	37	3.1	425	10	BF290883
9	37	3.1	514	10	BF282916
10	29	2.4	522	28	B56994
11	26	2.2	645	12	BI956079
12	26	2.2	750	13	BI719028
13	26	2.2	1069	13	BQ339574
14	25	2.1	156	29	CE072289
15	25	2.1	379	12	BM022612
16	25	2.1	401	10	BF564747
17	25	2.1	430	10	AW532942
18	25	2.1	441	9	AA292998
19	25	2.1	588	14	CF581582
20	25	2.1	927	13	BQ582829
21	25	2.1	1001	12	BI890470
22	25	2.1	1126	29	AG060944
23	25	2.1	1185	14	CF581581
24	24	2.0	272	9	AA048965
25	24	2.0	320	9	AI323408
26	24	2.0	347	10	BF523876
27	24	2.0	440	9	AI323053
28	24	2.0	442	28	BZ188256
29	24	2.0	646	14	CB576131
30	24	2.0	669	12	BI886947
31	24	2.0	808	14	CB960307
32	24	2.0	997	9	AL583668
33	24	2.0	1201	13	BX363285
34	23	1.9	129	14	CB245821
35	23	1.9	406	14	CAG23788
36	23	1.9	518	10	BB637563
37	23	1.9	518	14	CA698771
38	23	1.9	533	12	BJ063635
39	23	1.9	592	12	BM156494
40	23	1.9	672	12	BG546422
41	23	1.9	698	29	AG060332
42	23	1.9	717	12	BG489468
43	23	1.9	774	14	CK148319
44	23	1.9	790	12	BG716647
45	23	1.9	898	12	BG752658

#### ALIGNMENTS

RESULT 1  
BQ922461  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BQ922461  
AGENCOURT 8963457 Lupeki sciatic\_nerve Homo sapiens cdna clone  
IMAGE:6200329 5', mRNA sequence.  
BQ922461  
BQ922461.1 GI:22337492  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 904)

NIH-MGC http://mgs.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Dr. James R. Lupski

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: Llam13615 row: a column: 02

High quality sequence stop: 523.

FEATURES  
source  
1. .904  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6200329"  
/sex="male"  
/tissue\_type="sciatic nerve"  
/dev\_stage="adult, 70 yr"  
/lab\_host="DH10B"  
/clone\_lib="lupski sciatic nerve"  
/note="Vector: PCMV-SPORT6 (Life Technologies); Site 1:  
NotI; Site 2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCATCAGCGTCCG-3' and  
5'-GACATGTCAGTCGCGAGCGCCCTT(15)-3'. Size selected >  
1 kb for average insert length 1.87 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine) and is available through Life  
Technologies."

ORIGIN  
Query Match 31.1%; Score 376; DB 13; Length 904;  
Best Local Similarity 99.6%; Pred. No. 1.9e-132;  
Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 529 GTCAAGTGTGCGGACCCCTCCGCGCCCTGGGGCAAGCACTACTGATGCTCAAC 588  
Db 42 GTCAAGTGTGCGGACCCCTCCGCGCCCTGGGGCAAGCACTACTGATGCTCAAC 101  
Qy 589 CCCAACAGGAGTACACCTTCGCGGAGGGGTCTTCGCGCGCGCGCCCAAGACGCGCTCAGC 648  
Db 102 CCCAACAGGAGTACACCTTCGCGGAGGGGTCTTCGCGCGCGCGCCCAAGACGCGCTCAGC 161  
Qy 649 CACCGCGCGCGTCCCG 708  
Db 162 CACCGCGCGCGTCCCG 221  
Qy 709 GCG 768  
Db 222 GCG 281  
Qy 769 GAGGAGCG 828  
Db 282 GAGGAGCG 341  
Qy 829 CGCAAGCCCTTCGCGAGCGCTCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 888  
Db 342 CGCAAGCCCTTCGCGAGCGCGCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 401  
Qy 889 GCG 948  
Db 402 GCG 461  
Qy 949 AGGCGCTGTGCGCGCTGTGCGGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1006  
Db 462 AGGCGCTGTGCGCGCTGTGCGGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 519

RESULT 2  
BQ430047 982 bp mRNA linear EST 24-MAY-2002  
LOCUS AGNCOURT\_7916258 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:6010412  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ430047  
VERSION BQ430047.1 GI:21169123  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 982)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue procurement: DCTD/BTP/Gazdar  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LAM13198 row: g column: 21  
High quality sequence stop: 491.  
Location/Qualifiers  
1. .982  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6010412"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_68"  
/note="Organ: lung; Vector: PCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."

ORIGIN  
Query Match 12.1%; Score 146; DB 13; Length 982;  
Best Local Similarity 100.0%; Pred. No. 5.4e-45;  
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGAAGTTGGAGTGTTCCTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 265 ATGAAGTTGGAGTGTTCCTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 324  
Qy 61 GAGGCG 120  
Db 325 GAGGCG 384  
Qy 121 GCTCAGATGGGAGTGGCG 146  
Db 385 GCTCAGATGGGAGTGGCG 410

RESULT 3  
CE688385 246 bp DNA linear GSS 29-SEP-2003  
LOCUS tigr-gss-dog-17000368532929 Dog Library Canis familiaris genomic,  
DEFINITION Genomic survey sequence.  
ACCESSION CE688385  
VERSION CE688385.1 GI:37007455  
KEYWORDS GSS.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
REFERENCE 1 (bases 1 to 246)  
AUTHORS Kirkness, E.F., Baina, V., Halpern, A.L., Levy, S., Remington, K.,  
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and  
Venter, J.C.  
TITLE The dog genome: survey sequencing and comparative analysis  
JOURNAL Science 301 (5641), 1898-1903 (2003)  
MEDLINE 22875432  
PUBMED 14512627  
COMMENT The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: [ekirknes@tigr.org](mailto:ekirknes@tigr.org)

```

Class: shotgun.
Location/Qualifiers
1. 246
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 10.8%; Score 131; DB 29; Length 246;
Best Local Similarity 100.0%; Pred. NO. 2.5e-39;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 CTCACGACGCTTCGTCAGGTCGTCGCGACCCCTCGCGGCCCTCGGCGGCGGAGGACAC 573
DB 1 CTCACGACGCTTCGTCAGGTCGTCGCGACCCCTCGCGGCCCTCGGCGGCGGAGGACAC 60
QY 574 TACTGGATGTCACCCCAACAGCGAGTACACCTTCGCCGACGGGTCTTCGCCGCCGCC 633
DB 61 TACTGGATGTCACCCCAACAGCGAGTACACCTTCGCCGACGGGTCTTCGCCGCCGCC 120
QY 634 CGCAGCGCCT 644
DB 121 CGCAGCGCCT 131

RESULT 4
BM510453 406 bp mRNA linear EST 15-FEB-2002
LOCUS i341h03.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:5633525
DEFINITION 5, similar to TR:070220 070220 FORK HEAD TRANSCRIPTION FACTOR. 1,
mRNA sequence.
ACCESSION BM510453
VERSION BM510453.1 GI:186681596
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 406)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Secorce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other ESTs: i341h03.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioph.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pbluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seg primer: -40RP from Gibco
High quality sequence stop: 398.
Location/Qualifiers
1. 406
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cclone="IMAGE:5633525"
/tissue_type="insulinoma"

FEATURES
source
1. 406
/lab_host="Human insulinoma"
/clone_lib="Human insulinoma"
/notes="Organ: pancreas; Vector: pbluescript SK-; Site 1:
XhoI; Site 2: EcoRI. Constructed with lambda ZapII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pbluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

ORIGIN
Query Match 7.1%; Score 86; DB 12; Length 406;
Best Local Similarity 100.0%; Pred. No. 3.4e-22;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1068 CCCGCGCAAGCACTCCGAGCCCGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1127
DB 48 CCCGCGCAAGCACTCCGAGCCCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 107
QY 1128 GCTGCGCGCAGCCCTGCAGCGCGCT 1153
DB 108 GCTGCGCGCAGCCCTGCAGCGCGCT 133

RESULT 5
AI586081 398 bp mRNA linear EST 06-APR-1999
LOCUS v032c05.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
DEFINITION IMAGE:1209608 3', similar to gb:U36760 Mus musculus brain factor-1
(MOUSE);, mRNA sequence.
ACCESSION AI586081
VERSION AI586081.1 GI:4571978
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 398)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:645952
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 397.
Location/Qualifiers
1. 398
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/cclone="IMAGE:1209608"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse skin (#937313)"

```

/note="Organ: skin; Vector: pBluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

## ORIGIN

Query Match 6.4%; Score 77; DB 9; Length 398;  
Best Local Similarity 99.2%; Pred. No. 8.8e-19;  
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 526 TTCGTCAGGTCCTCGGACCCCTCGCGGCTTGGGCGAGGACACTACTGATGCTC 585  
|||||  
DB 44 TTGTCAGAGTCTCGCGACCCCTCGCGGCTTGGGCGAGGACACTACTGATGCTC 103  
|||||

QY 586 AACCCCAACAGCAGTACACCTTCGCCGACGGGGTCTTCGCCCGCCGCGCAAGCGCTC 645  
|||||  
DB 104 AACCCCAACAGCAGTACACCTTCGCCGACGGGGTCTTCGCCCGCCGCGCAAGCGCTC 163  
|||||

QY 646 AGCCACCG 653  
|||||  
DB 164 AGCCACCG 171  
|||||

## RESULT 6

AI169632/c  
LOCUS AI169632 559 bp mRNA linear EST 20-JAN-1999  
DEFINITION EST215519 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone  
RKBV46 3' end, mRNA sequence.

ACCESSION AI169632  
VERSION AI169632  
KEYWORDS EST.

SOURCE Rattus sp.  
ORGANISM Rattus sp.

## REFERENCE

1 (bases 1 to 559)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

## AUTHORS

Lee, N.H., Glöck, A., Chandra, I., Mason, T.M., Quackenbush, J.,  
Kerlavage, A.R. and Adams, M.D.

## TITLE

Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat  
Gene Index

## JOURNAL

Unpublished (1998)

## COMMENT

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

## FEATURES

## source

1..559  
Location/Qualifiers  
/organism="Rattus sp."  
/mol\_type="mRNA"  
/db\_xref="ATCC (inhost):2027517"  
/db\_xref="taxon:10118"  
/clone\_lib="RKIBV46"  
/note="Organ: kidney; Vector: pT73Pac; Site 1: EcoRI;  
Site 2: NotI"

## ORIGIN

Query Match 3.5%; Score 42; DB 9; Length 559;  
Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 TACACTTCGGCAGCGGGTCTTCGCCGCGCGCGCGCAAGCGC 642  
|||||

DB 548 TACACTTCGGCAGCGGGTCTTCGCCGCGCGCGCGCAAGCGC 507  
|||||

## RESULT 7

BI402320

LOCUS BI402320

DEFINITION MI-P-CP0-nwd-f-07-0-UI-s1 MI-P-CP0 Sus scrofa cDNA clone

MI-P-CP0-nwd-f-07-0-UI 3', mRNA sequence.

ACCESSION BI402320

VERSION BI402320.1

KEYWORDS GI:15181381

SOURCE EST.

ORGANISM Sus scrofa (pig)

REFERENCE Sus scrofa

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

TITLE Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

JOURNAL 1 (bases 1 to 232)

MEDLINE Ronaldo, M.F., Lennon, G. and Soares, M.B.

PUBMED Normalization and subtraction: two approaches to facilitate gene

COMMENT discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Tuggle CK

Molecular Genetics Laboratory, Department of Animal Science

Iowa State University

201 Kildee Hall, Ames, IA 50011-3150, USA

Tel: 5152944252

Fax: 5152942401

Email: cktuggle@iastate.edu

Oligo-dt track not found. Not 1 site shown in beginning of sequence

is likely internal to the message. cDNA Library Preparation: M.B.

Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,

University of Iowa Clone Distribution: clones will be available

through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=No.

## FEATURES

## source

1..232  
Location/Qualifiers  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/clone="MI-P-CP0-nwd-f-07-0-UI"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="MI-P-CP0"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: EcoRI. The MI-P-CP0  
library is derived from uterus. For a detailed description  
of the library from which this clone was derived, please  
visit our web site at http://pigest.genome.iastate.edu/.  
The procedure used to create this library has been  
previously described (Bonaldo, Lennon and Soares, Genome  
Research 6:791-806, 1996)  
TAG\_SEQ=None found"

## ORIGIN

Query Match 3.1%; Score 38; DB 12; Length 232;  
Best Local Similarity 100.0%; Pred. No. 0.00058;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 TCGCTCAACGACTGCTTGTCTCAAGTGTGCGCGACCC 548  
|||||

DB 105 TCGCTCAACGACTGCTTGTCTCAAGTGTGCGCGACCC 142  
|||||

## RESULT 8

BF290883/c

LOCUS BF290883

DEFINITION BF290883

ACCESSION BF290883

VERSION BF290883.1

KEYWORDS GI:11221953

SOURCE EST.

ORGANISM Rattus norvegicus (Norway rat)

REFERENCE Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

BF290883 425 bp mRNA linear EST 28-NOV-2000  
EST455474 Rat Gene Index, normalized rat, Rattus norvegicus cDNA  
Rattus norvegicus cDNA clone RGII165 3' sequence, mRNA sequence.

```

REFERENCE
AUTHORS      1 (bases 1 to 425)
Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Pertea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
TITLE        Generation of ESTs from Normalized Rat Embryo, Bento Soares
JOURNAL      Unpublished (2000)
COMMENT      Other ESTs: EST353025
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
FEATURES
source
Location/Qualifiers
1..425
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="RG11F65"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/notes="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RLI, RPL, RLU, REM, RMU,
RSP, RHE, RPC, RPN"
ORIGIN
Query Match      3.1%; Score 37; DB 10; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GCAAGTTCCTCCAGCTCTCGCCATCGACAGCATCCT 827
Db 343 GCAAGTTCCTCCAGCTCTCGCCATCGACAGCATCCT 307

RESULT 9
BF282916/c
LOCUS      BF282916      514 bp mRNA linear EST 28-NOV-2000
DEFINITION Rattus norvegicus cDNA clone RGIDU92 3' sequence, mRNA sequence.
VERSION    BF282916.1 GI:11213986
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE  1 (bases 1 to 514)
AUTHORS    Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Pertea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
TITLE      Generation of ESTs from Normalized Rat Embryo, Bento Soares
JOURNAL    Unpublished (2000)
COMMENT    Other ESTs: EST348098
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
FEATURES
source
Location/Qualifiers
1..514
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="RGIDU92"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"

```

```

/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/notes="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RLI, RPL, RLU, REM, RMU,
RSP, RHE, RPC, RPN"
ORIGIN
Query Match      3.1%; Score 37; DB 10; Length 514;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GCAAGTTCCTCCAGCTCTCGCCATCGACAGCATCCT 827
Db 356 GCAAGTTCCTCCAGCTCTCGCCATCGACAGCATCCT 320

RESULT 10
B56994
LOCUS      B56994      522 bp DNA linear GSS 20-JUN-1998
DEFINITION CIT-HSP-2008G17.TR CIT-HSP Homo sapiens genomic clone 2008G17,
genomic survey sequence.
ACCESSION  B56994
VERSION    B56994.1 GI:2611328
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 522)
AUTHORS    Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
Unpublished (1997)
Other GSSs: CIT-HSP-2008G17.TFB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamads@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7040419"
/db_xref="taxon:9606"
/clone="2008G17"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/notes="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"
ORIGIN
Query Match      2.4%; Score 29; DB 28; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 AGCCCCCTACTCGTACATCGCGCTCATC 381
Db 488 AGCCCCCTACTCGTACATCGCGCTCATC 516

RESULT 11
BI956079/c

```

LOCUS B1956079 645 bp mRNA linear EST 19-OCT-2001  
DEFINITION HVSMEM0025JU14f Hordeum vulgare green seedling EST library  
HVCNDA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA  
clone HVSMEM0025JU14f, mRNA sequence.  
ACCESSION B1956079  
VERSION B1956079  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
REFERENCE 1 (bases 1 to 645)  
AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D.,  
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.,  
Simmons, J., Oates, R. and Main, D.  
TITLE Development of a genetically and physically anchored EST resource  
for barley genomics: Blumeria infected Morex (compatible) seedling  
cDNA library  
JOURNAL Unpublished (2001)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total hg bases = 202  
Seq primer: AATTAACCTCCTACATAAGGG  
High quality sequence start: 32  
High quality sequence stop: 409.  
FEATURES  
source  
1..645  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Morex"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clones="HVSMEM0025JU14f"  
/tissue\_type="green seedling leaf"  
/lab\_host="TUC121"  
/clone\_lib="Hordeum vulgare green seedling EST library"  
HVCNDA0014 (Blumeria infected)  
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
XhoI; Morex (mla) plants were greenhouse grown in the R  
Wise lab at Iowa State University, Ames, IA; 7 day old  
green seedlings were infected with isolate 5874 of  
Blumeria graminis f. sp. hordei, and leaves were harvested  
24, 48 and 72 hr post-inoculation and snap frozen (Wise).  
In the TJ Close lab at the University of California,  
Riverside, total RNA was prepared from each sample pool,  
equal quantities of all three RNA pools were combined,  
poly(A) RNA was purified from the mixture, one primary  
unamplified cDNA library was made, and 1 million pfu were  
in vivo excised to give pBluescript SK(-) cDNA phagemids  
(Chin). Phagemids were plated and picked at the Clemson  
University Genomics Institute (CUGI) (Begum, Palmer,  
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
sequencing and sequence analysis were performed at CUGI  
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).  
The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders/Also  
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html")  
ORIGIN  
Query Match 2.2%; Score 26; DB 12; Length 645;

Best Local Similarity 100.0%; Pred. No. 22;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 699 CTTCCCGCGCGCGCGCGCGCGCGCGCGC 724  
|||||  
DB 598 CTTCCCGCGCGCGCGCGCGCGCGCGCGC 573  
|||||  
RESULT 12  
LOCUS BX719028 750 bp mRNA linear EST 18-NOV-2003  
DEFINITION BX719028 XGC-tadpole Silurana tropicalis cDNA clone ttpA063f13 5',  
mRNA sequence.  
ACCESSION BX719028  
VERSION BX719028  
KEYWORDS EST.  
SOURCE Silurana tropicalis (western clawed frog)  
ORGANISM Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
Xenopodinae; Silurana.  
REFERENCE 1 (bases 1 to 750)  
AUTHORS Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.  
TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)  
JOURNAL Unpublished (2003)  
COMMENT Contact: Croning MDR  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: TtpA063f13.plkSP6  
Sequencing primer: Sp6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Nigel Garrett.  
cDNA was oligo dt primed from sug of poly A+ RNA from tadpole  
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with  
EcoRI at the 5' end and NotI at the 3' end.  
Vector: pCS107; Site 1: EcoRI; Site 2: NotI  
Host: Escherichia coli DH10B.  
FEATURES  
source  
1..750  
/organism="Silurana tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TtpA063f13"  
/dev\_stage="tadpole (stage 35-40)"  
/lab\_host="E. coli DH10B"  
/clone\_lib="XGC-tadpole"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
was oligo dt primed from sug of poly A+ RNA from tadpole  
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107  
with EcoRI at the 5' end and NotI at the 3' end"  
ORIGIN  
Query Match 2.2%; Score 26; DB 13; Length 750;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 352 AAGCCCCCTACTCGTACATCGCGCT 377  
|||||  
DB 335 AAGCCCCCTACTCGTACATCGCGCT 360  
|||||  
RESULT 13  
LOCUS BQ939574 1069 bp mRNA linear EST 21-AUG-2002  
DEFINITION BQ939574 AGENCOURT 8931588 NTH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6484117  
5' mRNA sequence.  
ACCESSION BQ939574  
VERSION BQ939574.1 GI:22355052  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1069)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: Rubin Laboratory  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM269 row: m column: 14  
High quality sequence spot: 228.  
Location/Qualifiers  
1..1069  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:648417"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_40"  
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed in  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

Query Match 2.2%; Score 26; DB 13; Length 1069;  
Best Local Similarity 100.0%; Pred.No. 23;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 707 CGCGCCCGCGCGCGCGCGCGCGCGC 732  
DB 880 CGCGCCCGCGCGCGCGCGCGCGCGC 855

CE072289 156 bp DNA linear GSS 24-SEP-2001  
tigr-gss-dog-17000323549470 Dog Library Canis familiaris genomic,  
genomic survey sequence.

CE072289  
CE072289.1 GI:35135052  
GSS.  
Canis familiaris (dog)  
Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
1 (bases 1 to 156)  
Kirkness,E.F., Batina,V., Halpern,A.L., Levy,S., Remington,K.,  
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
Venter,J.C.  
The dog genome: survey sequencing and comparative analysis  
Science 301 (3641), 1898-1903 (2003)

CE072289/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

pancreas, newborn pancreas, adult pancreas, and adult islets of Langerhans were separately constructed using Superscript Plasmid Library kit (Life Technologies). cDNA was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:731-806, 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

ORIGIN

Query Match 2.1%; Score 25; DB 12; Length 379;  
Best Local Similarity 100.0%; Pred. NO. 52;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 CCTACTCGTACATCGCGTCTATCG 382  
|||||  
DB 266 CCTACTCGTACATCGCGTCTATCG 242  
|||||

Search completed: March 31, 2004, 04:42:53  
Job time : 2986 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 29, 2004, 11:52:41 ; Search time 64.5 Seconds  
(without alignment)  
10592.246 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2189

Sequence: 1 atgaagtggagggtgtcgt.....tggagagcctcctagcttga 1209

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool\_p/US10087080/runat 29032004 114525 6520/app\_query.fasta\_1.1351  
-DB=A\_Geneseq\_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087080 @CGN 1.1.81 @runat 29032004 114525 6520 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Databases :

A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003s:\*  
7: Geneseq2003ss:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2123	97.0	402	5 ABP54691	Abp54691 Metastati
2	1314.5	60.1	439	7 ADD48790	Add48790 Rat Prote
3	1314.5	60.1	439	7 ADD47621	Add47621 Rat Prote
4	517	23.6	465	7 ADB75312	Adb75312 Prostate
5	467	21.3	478	6 ABR63544	Abri63544 Danio rer
6	467	21.3	478	7 ADD48650	Add48650 Human Pro
7	467	21.3	478	7 ADD46993	Add46993 Human Pro
8	467	21.3	478	7 ADD46997	Add46997 Human Pro
9	466	21.3	464	6 ABR63545	Abri63545 Danio rer
10	445	20.3	396	6 ABR63546	Abri63546 Danio rer

11	436.5	19.9	409	6	ABU56559	Abu56559 Lung canc
12	431.5	19.7	553	2	AAV43260	Aay43260 Human FKX
13	431.5	19.7	553	2	AAV55721	Aay55721 Human for
14	429.5	19.6	553	2	AAV41277	Aay41277 Human FRE
15	415.5	19.0	376	5	ABJ01138	Abj01138 Ovary cel
16	402.5	18.4	478	4	ABG08496	Abg08496 Novel hum
17	402	18.4	371	6	ABR63547	Abri63547 Danio rer
18	394.5	18.0	330	4	ABG1327	Abg1327 Human tra
19	393	18.0	465	4	ABG03158	Abg03158 Novel hum
20	387.5	17.7	501	2	AAV01097	Aay01097 FREAC11 p
21	387	17.7	473	5	AAU74806	Aau74806 Human hep
22	387	17.7	473	5	AAU74807	Aau74807 Human hep
23	387	17.7	473	6	ABR47488	Abri47488 Breast ca
24	384.5	17.6	501	4	AAAB2851	Aab2851 Human win
25	384.5	17.6	501	5	AAU79816	Aau79816 Human tra
26	384.5	17.6	501	6	ADA50747	Ada50747 Human FOX
27	384	17.5	494	5	AAU79818	Aau79818 Mouse mes
28	373	17.0	19938	6	ABP76680	Abp76680 Streptomy
29	366.5	16.3	19938	6	ABP76680	Abp76680 Streptomy
30	366	16.7	371	6	ABR63543	Abri63543 Danio rer
31	366	16.7	428	4	ABG08497	Abg08497 Novel hum
32	365	16.7	19938	6	ABP76681	Abp76681 Streptomy
33	360	16.4	19938	6	ABP76681	Abp76681 Streptomy
34	354.5	16.2	456	4	ABB60077	Abb60077 Drosophi
35	347.5	15.9	19938	6	ABP76682	Abp76682 Streptomy
36	345	15.8	480	2	ABR44551	Abri44551 Brain fac
37	342.5	15.6	19938	6	ABP76679	Abp76679 Streptomy
38	341	15.6	481	5	ABBS7076	Abbs7076 Mouse isc
39	338.5	15.5	477	6	ABP72344	Abp72344 Brain fac
40	338	15.4	509	5	ABG95654	Abg95654 Human nuc
41	336	15.0	19938	6	ABP76681	Abp76681 Streptomy
42	333.5	15.2	182	5	ABP43874	Abp43874 Transcrip
43	332.5	15.2	19938	6	ABP76678	Abp76678 Streptomy
44	329	15.0	544	2	AAW90247	Aaw90247 Human PAS
45	326.5	14.9	451	4	ABB59467	Abb59467 Drosophi

ALIGNMENTS

RESULT 1

ABP54691

ID ABP54691 standard; protein; 402 AA.

XX

AC ABP54691;

XX

DT 30-DEC-2002 (first entry)

XX Metastatic colorectal cancer-associated polypeptide.

DE Colorectal cancer; metastasis; differential expression; cytostatic;

XX Colorectal cancer; metastasis; differential expression; cytostatic;

KW diagnosis; gene therapy; vaccine.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT Misc-difference 387

FT /note= "encoded by CGN"

XX

PN WO200268677-A2.

XX

PD 06-SEP-2002.

XX

PF 27-FEB-2002; 2002WO-US006001.

XX

PR 27-FEB-2001; 2001US-0272206P.

PR 02-APR-2001; 2001US-0281149P.

PR 17-APR-2001; 2001US-0284555P.

XX

PA (E0SB-) EOS BIOTECHNOLOGY INC

PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX

PI Mack DH, Markowitz SD;

XX



DR GENBANK; Q63244.  
XX New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
PT  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 439 AA;  
  
Alignment Scores:  
Pred. No.: 1,096-74 Length: 439  
Score: 1314.50 Matches: 278  
Percent Similarity: 74.30% Conservative: 14  
Best Local Similarity: 70.74% Mismatches: 86  
Query Match: 60.05% Indels: 15  
DB: 7 Gaps: 6  
  
US-10-087-080-31 (1-1209) x ADD48790 (1-439)  
  
QY 49 GGCAGTACCTGGAGGGCGCGGGCGGACGACGCGCGTCCCGCTGTCGGCGCGGGA 108  
DB 2 GlySerAspLeuGluGlyAlaGlySerSerAspValProSerProLeuSerAlaAlaGly 21  
QY 109 GACGACTCCCTGGCTCAGATGGGACTGCGCGCC---AAGCCGTCCGCGGCGCGCG 165  
DB 22 AspAspSerLeuGlySerAspGlyAspCysAlaAlaAsnSerProAlaAlaGlySerGly 41  
QY 166 GCCAGAGATACGAG---GGCAGCGGCAACAGAGTGGGAGGGCGCGCGCGCGGAG 222  
DB 42 AlaGlyAspLeuGluGlyGlyGlyGluArgAsnSerSerGlyGlyAlaSerThrGln 61  
QY 223 GAGCGGATCCCGGACGAGCTGCTCAGCGGTGTGTGGCGAGGGCGCG---GAGGCC 276  
DB 62 Asp-----AspProGluValThrAspGlySerArgThrGlnAla 74  
QY 277 GGGCGCGGGGCGCAGGCGCGGGCGGGCGGAGCGGAGGTGCACGACGAGCA 336  
DB 75 SerProValGlyProCysAlaGlySerValGlyGlyGlyGluGlyAlaArgSerLysPro 94  
QY 337 TATAGCGGGCGGCCCAAGCCCCCTTACTCGTACATCGCGCTCATCGCCATCGCCATCCGC 396  
DB 95 TyrThrArgArgProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArg 114  
QY 397 GACTCGGGCGGGCGCGCTTGAAGTGGCGGAGATCAACGAGTACCTCATCGGCAAGTTC 456

DB 115 AspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPhe 134  
QY 457 CCCTTTTTCGCGCGAGCTACACGGCTGGCGCACTCCGTGGCCACACCACTTCGCTC 516  
DB 135 ProPhePheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeu 154  
QY 517 AACGACTGCTTCGTCAGGTGCTGCGGACCCCTCGGGCCCTCGGGCAAGGCACTAC 576  
DB 155 AspAspCysPheValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyr 174  
QY 577 TGGATGCTCAACCCCAACACGAGCAGTACACTTCGCGACGGGGTCTTCGCGCCGCCCG 636  
DB 175 TrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgGly 194  
QY 637 AACGGCTAGCCACCGCGCGCGCGTCCCGCGCCCGCGCGCGCGCGCGCGCGCGCG 696  
DB 195 LysArgLeuSerHisArgThrThrValSerAlaSerGlyTyrGlyGlySerProPro 214  
QY 697 GGCTCCCG 756  
DB 215 GlyProAlaGlyThrProGlnProAlaProThrAlaGlySerSerProIleAlaArgSer 234  
QY 757 CCG 816  
DB 235 ProAlaArgGlnGluGlySerSerProAlaSerLysPheSerSerSerPheAlaIle 254  
QY 817 GACAGCATCTCGCGAAGCCCTTCGCGCGCTCAGGACACACGCGCGCGCGCGCG 876  
DB 255 AspSerIleLeuSerLysProPheArgSerArgArg-AspGlyThrArgLeuTrpGlyCy 274  
QY 877 ACCTTCAGTGGCG 936  
DB 274 sSerTyrProGlyAlaLeuLeuProAlaArgArgCysAlaProIleProArgSerPhePr 294  
QY 937 GCGGCG 996  
DB 294 oArgArgProAlaValProCysArgSerValLeuThrAlaArg-ArgAlaHisAla 314  
QY 997 CTGGCG 1056  
DB 314 laGlyValAlaArgGlyArgGlyAlaAla---ArgGlyAlaProValArgGlyAlaProL 333  
QY 1057 CCG 1113  
DB 333 euHisArgGlyProSerGlnAlaIleSerArgSerGlyAspArgArgGlyAlaProV 353  
QY 1114 TACTGCCCCCTGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1173  
DB 353 alLeuProProThrAlaAlaHisGlyProAlaGlyArgGlyLeuArgSerGlySerA 373  
QY 1174 CACTGTGTACTCCGCGTGGAGCGCTCCTAGCTTG 1208  
DB 373 laProValLeuProGlyGlyAspAlaAlaSerLeu 384  
  
RESULT 3  
ADD47621  
ID ADD47621 standard; protein; 439 AA.  
XX  
AC ADD47621;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Rat Protein Q63244, SEQ ID NO 13317.  
XX  
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
XX  
OS Rattus norvegicus.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX





RESULT 5  
ABR63544  
ID ABR63544 standard; protein; 478 AA.  
XX  
AC ABR63544;  
XX  
DT 18-SEP-2003 (first entry)  
XX  
DE Danio rerio foxd3 homologue human HPH2.  
XX  
DE Craniofacial malformation; variant foxd3; osteopathic; cardiant;  
KW cytotatic; auditory; neuroprotective; cleft palate syndrome;  
KW congenital heart disease; cancer; human; HPH2.  
XX  
OS Homo sapiens.  
XX  
PN WO2003048196-A2.  
XX  
PD 12-JUN-2003.  
XX  
PF 09-DEC-2002; 2002WO-EP013936.  
XX  
PR 07-DEC-2001; 2001EP-00129205.  
XX  
PA (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.  
XX  
PI Knapik E, Sachdev S;  
XX  
DR WPI; 2003-513737/48.  
XX  
PT New foxd3 polynucleotide and their (ant)agonists, useful for diagnosing  
PT and treating craniofacial malformations, such as cleft palate syndrome,  
PT loss of middle ear ossicles, congenital heart disease, cancer and mental  
PT retardation.  
XX  
PS Claim 34; Page 64-66; 70pp; English.  
XX  
CC The present invention provides the protein and coding sequences of Danio  
CC rario foxd3. These can be used in the diagnosis and treatment of  
CC craniofacial malformations, such as cleft palate syndrome, loss of middle  
CC ear ossicles, congenital heart disease, Nishikawa-Kuroki syndrome, cancer  
CC and mental retardation. The present sequence is a homologue of the  
CC zebrafish foxd3 protein, and is human HPH2  
XX  
SQ Sequence 478 AA;  
  
Alignment Scores:  
Pred. No.: 2,09e-21 Length: 478  
Score: 467.00 Matches: 161  
Percent Similarity: 39.88% Conservative: 36  
Best Local Similarity: 32.59% Mismatches: 164  
Query Match: 21.33% Indels: 134  
DB: 6 Gaps: 17  
  
US-10-087-080-31 (1-1209) x ABR63544 (1-478)  
QY 37 GGGGACAAAGCAGGCGAGTACCTGGAGGC-----CGGGCGGCGAGCGAGCGCGC 87  
Db 5 GlyGlySerAlaSerAspMetSerGlyGlnThrValLeuThrAlaGluAspValAsp 24  
QY 88 TCCCGCTGTCGGCGGCGGAGACGACTCCCTGGGCTCAGATGGGACTGCGCGGCCAAG 147  
Db 25 IleAspValValGlyGluGlyAspGlyLeuGluGlyAspSerAspAlaGlyCys 44  
QY 148 CCGTCCGCGGCGGCGGCGGCGGAGATACGCGGCGGAGCGGCGGAGAGAGTG----- 199  
Db 45 AspSerProAlaGly-ProProGluLeuArgLeuAspGluAlaAspGluValProAla 64  
QY 200 -----CGGA 204  
Db 64 aAlaProHisGlyGlnProGlnProHisGlnGlnProLeuThrLeuProLysG 84

QY 205 GCGGCGCGGCGCGGCGA-----GGAGGCGAT-----CCGGGCGAGCAGCTGCT 246  
Db 84 uAla-AlaGlyAlaGlyAlaGlyProGlyGlyAspValGlyAlaProGluAlaAspGlyC 104  
QY 247 GCAGCGTGTGTGGCGGAGGCGC-----GGAGCGCGGCGCGGCGGCGCGCGGCG 300  
Db 104 ySLysGlyGlyValGlyGlyGluGluGlyAlaSerGlyGly-GlyProGlyAlaGly 123  
QY 301 GCGCGGCGGAGCGGCGAGGTGTACGACGAGCAAGCATATACGCGGCGGCGGCGGCGG 360  
Db 124 SerGlySerAlaGlyLeuAlaProSerLysProLysAsnSerLeuValLysProPro 143  
QY 361 TACTCGTACATCGGCTCATCGCATCGCATCGGAGCTCGGCGGCGGCGGCGGCTTACG 420  
Db 144 TyrSerTyrIleAlaLeuIleThrMetAlaIleLeuGlnSerProGlnLysLeuThr 163  
QY 421 CTGCGGAGATCAACAGATCTCATGGCAAGTTCCTTTTTCGCGGCGGAGCTACAG 480  
Db 164 LeuSerGlyIleCysGluPheIleSerAsnArgPheProTyrTyrArgGluLysPhePro 183  
QY 481 GGCTGGCGCAACTCCGTGGCGCCACACTTTCCTCAACGACTGCTTCGTCAAGGTCTG 540  
Db 184 AlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPheValLysIlePro 203  
QY 541 CGGACCCCTCGCGGCGCTGGGCAAGCAACTACTGTGATCTCAACCCCAACAGCGAG 600  
Db 204 ArgGluProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLeuAspProGlnSerGlu 222  
QY 601 TACACCTTGGCGGAGCGGCTTCCGCGCGCGGCGGAGCGGCTCAGCCACCGCGCGCC- 659  
Db 223 AspMetPheAspAsnGlySerPheLeuArgArgLysArgPheLysArgHisGlnGln 242  
QY 660 -----GGTCCCGCGCGCGCGCGCT 677  
Db 243 GluHisLeuArgGluGlnThrAlaLeuMetMetGlnSerPheGlyAlaTyrSerLeuAla 262  
QY 678 GCGGCGGAGAGCGCGCGGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737  
Db 263 AlaAlaAlaGlyAlaAlaGlyProTyrGlyArgPro-----TyrGlyLeuHisProAla 280  
QY 738 CTCGCCCGGATGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAAGTT 797  
Db 281 AlaAlaAlaGlyAlaTyrSerHisProAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300  
QY 798 CTCAGGCTCTTCGCCCATTCGACAGATCTTCGGCAAGCCCTTCGCGAGCGCTCG- 851  
Db 301 LeuGlnTyrProTyrAlaLeuProProValAlaProValLeuPro---ProAlaValPro 319  
QY 852 -----CCTCAGGAGACAGCGCGCGCGCGCGCGCTTCAGTGGGCGCGCG- 896  
Db 320 LeuLeuProSerGlyGluLeuGlyArgLysAlaAlaAlaAlaAlaAlaAlaAlaAla 339  
QY 897 -----GCCCTCGCGCGCGCTCGCGCGCGCGCGCGT----- 920  
Db 340 GlyLeuGlnLeuGlnLeuAsnSerLeuGlyAlaAlaAlaAlaAlaAlaAlaGlyThrAlaGly 359  
QY 921 -----CCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 944  
Db 360 AlaAlaGlyThrAlaSerLeuIleLysSerGluProSerAlaArgProSerPheSer 379  
QY 945 CTGCAG-----GGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 989  
Db 380 IleGluAsnIleIleGlyGlyGlyProAlaAlaPro-----GlyGlySerAla 395  
QY 990 GCGCGGCTGGCG 1049  
Db 396 ValGlyAlaGlyAlaAlaGlyGlyThrGlyGly----- 406  
QY 1050 ACCTCTCCCG 1106  
Db 407 ---SerGlyGlyGlySerThrAlaGlnSerPheLeuArgProProGlyThrValGlnSer 425  
QY 1107 -----GCA 1109

Db 426 AlaAlaLeuMetAlaThrHisGlnProLeuSerLeuSerArgThrThrAlaThrIleAla 445  
QY 1110 CQTGFACTS-----CCTGCGCGCTGCGCGCAGC 1139  
Db 446 ProIleLeuSerValProLeuSerGlyGlnPheLeuGlnProAlaAlaSerAlaAlaAla 465  
QY 1140 CQTGAGCGCGCTAGTCGCGCTGCGCGCGCA 1175  
Db 466 AlaAlaAlaAlaAlaAlaGlnAlaLysTrpProAla 477  
RESULT 6  
ADD48650  
ID ADD48650 standard; protein; 478 AA.  
XX ADD48650;  
AC  
XX  
XX 29-JAN-2004 (first entry)  
XX  
XX Human Protein NP\_036315, SEQ ID NO 14356.  
XX  
XX Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; Chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
XX Homo sapiens.  
XX  
XX WO2003016475-A2.  
XX  
XX  
XX 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
XX 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEHO ) GEN HOSPITAL CORP.  
XX (FARB ) BAYER AG.  
XX  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX  
XX WPI; 2003-268312/26.  
XX GENBANK; NP\_036315.

XX New composition comprising two or more isolated polypeptides, useful for  
XX preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
XX derivative or allelic variation of the nucleic acid sequence. Also  
XX claimed are a vector comprising the novel polynucleotide, a host cell  
XX comprising the vector, a method for identifying a nucleotide sequence  
XX which is differentially regulated in an animal subjected to pain and a  
XX kit to perform the method, an array, a method for identifying an agent  
XX that increases or decreases the expression of the polynucleotide sequence  
XX that is differentially expressed in neuronal tissue of a first animal  
XX subjected to pain, a method for identifying a compound which regulates  
XX the expression of a polynucleotide sequence which is differentially  
XX expressed in an animal subjected to pain, a method for identifying a  
XX compound that regulates the activity of one or more of the  
XX polynucleotides, a method for producing a pharmaceutical composition, a  
XX method for identifying a compound or small molecule that regulates the  
XX activity in an animal of one or more of the polypeptides given in the  
XX specification, a method for identifying a compound useful in treating  
XX pain and a pharmaceutical composition comprising the one or more  
XX polypeptides or their antibodies. The polynucleotide or the compound that  
XX modulates its activity is useful for preparing a medicament for treating  
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
XX therapy). The sequence presented is a human protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 478 AA;  
Alignment Scores: 2.09e-21 Length: 478  
Pred. No.: 467.00 Matches: 161  
Score: 39.88% Conservative: 36  
Percent Similarity: 32.59% Mismatches: 164  
Best Local Similarity: 21.33% Indels: 134  
Query Match: 7 Gaps: 17  
DB:  
US-10-087-080-31 (1-1209) x ADD48650 (1-478)  
QY 37 GGGGACAGCAGGCGAGTGTGAGCTGGAGGGC-----GCGGCGCGACGCGCGCGC 87  
Db 5 GlyGlyGlySerAlaSerAspMetSerGlyGlnThrValLeuThrAlaGluAspValAsp 24  
QY 88 TCCCGCGCTGTCGCGCGGAGAGACGACTCCCTCGGCTCAGATGGGACTCGCGCGCCAG 147  
Db 25 IleAspValValGlyGlyGlyAspGlyLeuGluGlyLysAspSerAspAlaGlyCys 44  
QY 148 CCGTCCCGCGCGCGCGCGCGAGATACGAGGCGCGCGAGACAGAGTG-----199  
Db 45 AspSerProAlaGly-ProProGluLeuArgLeuAspGluAlaAspGluValProAla 64  
QY 200 -----CGGA 204  
Db 64 aAlaProHisGlyGlnProGlnProHisGlnGlnProLeuThrLeuProLysG 84  
QY 205 GCGGCGCGCGCGCGG-----GGAGCGAT-----CCGCGGAGAGCTGCT 246  
Db 84 uAla-AlaGlyAlaGlyAlaGlyProGlyGlyAspValGlyAlaProGluAlaAspGlyC 104  
QY 247 GCAGCGTGTGCGCGGAGGCGC-----GGAGCGCGCGCGCGCGCGCGCGCG 300  
Db 104 yLysGlyGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 123  
QY 301 GCGCGCGGAGCGCGGAGGCTGACGAGCAAGCATATACGCGCGCGCGCGCGCGCGCG 360  
Db 124 SerGlySerAlaGlyLeuAlaProSerLysProLysAsnSerLeuValLysProPro 143  
QY 361 TACTCGTACATCGGCTCATCGCATCGCATCGCATCGCGCGCGCGCGCGCGCGCGCG 420  
Db 144 TyrSerTyrIleAlaLeuIleThrMetAlaIleLeuGlnSerProGlnLysLeuThr 163  
QY 421 CTGCGGAGATCAACGAGTACCTCATCGGCAAGTTCCCTTTTTCGCGCGCGCTACAG 480  
Db 164 LeuSerGlyIleCysGluPheIleSerAsnArgPheProTyrTyrArgGluLysPhePro 183  
QY 481 GGTGCGCACTCGTGGCGCCACCACTTCGCTCAACGACTGCTTCTCGTCAAGTGCTG 540  
Db 184 AlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPheValLysIlePro 203  
QY 541 CGCGACCCCTCGCGCGCGCTGGGCAAGGCAACTACTGATGTCTCAACCCCAACAGCGAG 600  
Db 204 ArgGluProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLeuAspProGlnSerGlu 222  
QY 601 TACACTTCGCGAGCGGGCTTTCGCGCGCGCGCGCGAGCGCTCAGCCACCGCGCGC 659  
Db 223 AspMetPheAspAanGlySerPheLeuArgArgArgLysArgPheLysArgHisGlnGln 242  
QY 660 -----GTCCTCCCGCGCGCGCGCT 677  
Db 243 GluHisLeuArgGluGlnThrAlaLeuMetMetGlnSerPheGlyAlaTyrSerLeuAla 262  
QY 678 GCGGCGCGAGAGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737  
Db 263 AlaAlaAlaGlyAlaAlaGlyProTyrGlyArgPro-----TyrGlyLeuHisProAla 280





Alignment Scores:			
Pred. No.:	2,09e-21	Length:	478
Score:	467.00	Matches:	161
Percent Similarity:	39.88%	Conservative:	36
Best Local Similarity:	32.59%	Mismatches:	164
Query Match:	21.33%	Indels:	134
DB:	7	Gaps:	17
US-10-087-080-31 (1-1209) x ADD46997 (1-478)			
QY	37	GGGAGCAAGCAGGCGAGTACCTGGAGGCG	-----GCGGGCGGACGACGCGCGC 87
DB	5	GlyGlySerAlaSerMetSerGlyGlnThrValLeuThrAlaGluAspValAsp 24	
QY	88	TCCCCCTGTGCGCGCGGAGACGACTCCCTGGGTTCAGATGGGACTGCGCGCCCAAG 147	
DB	25	IleAspValValGlyGluGlyAspGlyLeuGluGluLysAspSerAspAlaGlyCys 44	
QY	148	CCCTCCGCGCGCGCGCGCCAGATACGACGGCGGACGAGAGTGTG----- 199	
DB	45	AspSerProAlaGly-ProGluLeuLeuArgLeuAspGluAlaAspGluValProAla 64	
QY	200	-----CGGGA 204	
DB	64	aAlaProHisGlyGlnProGlnProHisGlnGlnProLeuThrLeuProLysGln 84	
QY	205	GGCGGCGCGCGCGGGA-----GGAGCGAT-----CCGCGACAGCTGCT 246	
DB	84	uAla-AlaGlyAlaGlyAlaGlyProGlyGlyAspValGlyAlaProGluAlaAspGlyC 104	
QY	247	GCAGCGTGTGTGCGGAGGCGCG-----GGAGCGCGCGCGCGCGCGCGCGCGC 300	
DB	104	ysLysGlyGlyValGlyGlyGluGluGlyAlaSerGlyGly-GlyProGlyAlaGly 123	
QY	301	GGCGCGGAGCGCGGAGGCGGTCACGACGACGACGACGACGACGACGACGACGACG 360	
DB	124	SerGlySerAlaGlyGlyLeuAlaProSerLysProLysAsnSerLeuValLysProPro 143	
QY	361	TACTCGTACATCGCTCATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGC 420	
DB	144	TySerTrpIleAlaLeuIleThrMetAlaIleLeuGlnSerProGlnLysLysLeuThr 163	
QY	421	CTGCGGAGATCAACGAGTACCTCATGGGAGTTCCTCTCAAGCTGCTTCCTCAAGTGCTG 540	
DB	184	AlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPheValLysIlePro 203	
QY	541	CGCGACCCCTCGCGCGCGCTGGGCAAGGACAACTACTGGATGCTCAACCCCAACAGCGAG 600	
DB	204	ArgGluProGlyAsnPro---GlyLysGlyAsnTy-TrpThrLeuAspProGlnSerGlu 222	
QY	601	TACACCTTCGCGCGCGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC- 659	
DB	223	AspMetPheAspAsnGlySerPheLeuArgArgArgLysArgPheLysArgHisGlnGln 242	
QY	660	-----GGTCCCGCGCGCGCGGCT 677	
DB	243	GluHisLeuArgGluGlnThrAlaLeuMetMetGlnSerPheGlyAlaTySerLeuAla 262	
QY	678	GCGCGCGCGAGAGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 737	
DB	263	AlaAlaAlaGlyAlaAlaGlyProTyGlyArgPro-----TyGlyLeuHisProAla 280	
QY	738	CTGCGCGCGATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAAGTT 797	
DB	281	AlaAlaAlaGlyAlaTySerHisProAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300	
QY	798	CTCAGCTCTTCGCCCATGACAGCATCTCTGCGCAAGCCCTTCGCGAGCGCTCG----- 851	
DB	301	LeuGlnTyProTyAlaLeuProProValAlaProValLeuPro---ProAlaValPro 319	

RESULT 9

ID	ABR63545	standard; protein; 464 AA.
XX	ABR63545;	
AC	ABR63545;	
DT	18-SEP-2003 (first entry)	
DE	Danio rerio foxd3 homologue murine HFX2.	
KW	Craniofacial malformation; variant foxd3; osteopathic; cardiac;	
KW	cytostatic; auditory; neuroprotective; cleft palate syndrome;	
KW	congenital heart disease; cancer; mouse; HFX2.	
OS	Mus musculus.	
PN	WO2003048196-A2.	
PD	12-JUN-2003.	
PF	09-DEC-2002; 2002MO-EP013936.	
PR	07-DEC-2001; 2001EP-00129205.	
XX	(GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.	
PI	Knapik E, Sachdev S;	
XX	WPI; 2003-513737/48.	
XX	New foxd3 polynucleotide and their (ant)agonists, useful for diagnosing	
PT	and treating craniofacial malformations, such as cleft palate syndrome,	
PT	loss of middle ear ossicles, congenital heart disease, cancer and mental	
XX	retardation.	
PS	Claim 34; Page 66-67; 70pp; English.	

XX The present invention provides the protein and coding sequences of Danio  
 CC rerio foxd3. These can be used in the diagnosis and treatment of  
 CC craniofacial malformations, such as cleft palate syndrome, loss of middle  
 CC ear ossicles, congenital heart disease, Niikawa-Kuroki syndrome, cancer  
 CC and mental retardation. The present sequence is a homologue of the  
 CC zebrafish foxd3 protein, and is murine HFH2

XX SQ Sequence 464 AA;

Alignment Scores:  
 Pred. No.: 2,426-21 Length: 464  
 Score: 466.00 Matches: 154  
 Percent Similarity: 39.25% Conservative: 34  
 Best Local Similarity: 32.15% Mismatches: 173  
 Query Match: 21.29% Indels: 118  
 DB: 6 Gaps: 16

US-10-087-080-31 (1-1209) x ABR63545 (1-464)

QY 37 GGGACACAGGGGCGAGTACTGAGGCG-----GCGGGGCGAGCGAGCGCGG 87  
 DB 5 GlySerGlySerAlaSerAspMetSerGlyGlnThrValLeuThrAlaGluAspValAsp 24  
 QY 88 TCCCGCTGTCCGGCGGAGAGACTCCCTG-----GGCTCAGATGGGACTGC 138  
 DB 25 IleAspValValGlyGluAspGlyLeuGluGluLysAspSerAspAlaGlyCys 44  
 QY 139 -----GCGGCAAGCGC-----TCC 153  
 DB 45 SerProAlaGlyProProAspLeuArgLeuAspGluAlaAspGluGlyProValSer 64  
 QY 154 GGGGGCGGGCGGAGATACGAGCGGCGGCGGAGCGGAGCGGCGGCGGCGG 213  
 DB 65 AlaHisGlyGlnSerGlnProGlnAlaLeuAlaLeuProThrGluAlaThrGlyPro 84  
 QY 214 GCGCGGAGAGCGGATCCCGGAGCGAGCTGTGTCAGCGGTGTGCGGAGGCGCGGAG 273  
 DB 85 GlyAsnAspThrGlyAlaProGluAlaAspGlyCysLys--GlyGlyGluAspAlaVal 103  
 QY 274 GCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 333  
 DB 104 ThrGlyGlyGlyProGlyAlaGlySerGlyAlaThrGlyGlyLeuThrProAsnLys 123  
 QY 334 CCATATACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 393  
 DB 124 ProLysAsnSerLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIle 143  
 QY 394 CCGGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453  
 DB 144 LeuGlnSerProGlnLysLysLeuThrLeuSerGlyIleCysGluPheIleSerAsnArg 163  
 QY 454 TTCCCTTTTTCGGGCGGAGTACACGGGCTCGGCGCAACTCGGCGGCGGCGGCGGCGG 513  
 DB 164 PheProTyrTyrArgGluLysPheProAlaThrGlnAsnSerIleArgHisAsnLeuSer 183  
 QY 514 CTCACAGACTGTTCGTCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573  
 DB 184 LeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro--GlyLysGlyAsn 202  
 QY 574 TACTGGATGCTCAACCCCAACAGCGAGTACCTTCGCGGCGGCGGCGGCGGCGGCGGCGG 633  
 DB 203 TyrTrpThrLeuAspProGlnSerGluAspMetPheAspAsnGlySerPheLeuArgArg 222  
 QY 634 CGCAAGCGC----- 642  
 DB 223 ArgLysArgPheLysArgHisGlnGlnHisLeuArgGluGlnThrAlaLeuMetMet 242  
 QY 643 -----CTCAGCCACCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 681  
 DB 243 GlnSerPheGlyAlaTyrSerLeuAlaAlaAlaGlyAlaGlyProTyrGlyLeuHis 262  
 QY 682 CCGGAGGAGCG 741

DB 263 ProAlaAlaAlaAlaGly-----AlaTyrSerHisProAlaAlaAlaAlaAlaAla 280  
 QY 742 CCGCGCATGCGCTCGCCCG-----CCCGCCAG 768  
 DB 281 -GlyCysCysGlyAlaProValProValArgAlaThrThrArgGlySerArgAlaAla 300  
 QY 769 GAGGAGCGCCAGCG-----CCGCGGCAAGTTCCTCCAGCTCCTTC 810  
 DB 300 aArgSerAlaProAlaAlaLeuGlyArgAlaGlyProGlnSerGlyArg-LeuArgLeuA 320  
 QY 811 GCCATCGACAGCATCTCGCCCAAGCCCTCCGAGCG-----TCGCCTCAGG 858  
 DB 320 laalaArgSerGluProAlaAlaThrAlaGlnHisProGlyArgArgArgGlySerArg 340  
 QY 859 GACACGCGCGCGGAGCGGCTTCAAGTGGGCGC-----CGCGCGCTCCCGCGCGCTG 912  
 DB 340 lylHisGlyGlyArgGlyGlyHisValAlaHisGlnValArgAlaGlnCysAlaAlaV 360  
 QY 913 CCGCGTTCCTCCGCGCTCTCCCGGCGGCGCTGAGGCGCTGCTGCGCTCTGCGCG 972  
 DB 360 alValGlnHisArgGluHisArgGlyIleArgAlaProGlyGlySer----- 377  
 QY 973 TACGCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1032  
 DB 378 -----AlaGlyGlyGlySerGlyGlyAlaGlyGlyGlyGly----- 392  
 QY 1033 CCGCGCTCTGCTGTCACCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1092  
 DB 393 -----SerGlyGlyGlyGlyAlaGlnSerPheLeuArgProp 406  
 QY 1093 GCGGCGCGGCGGCG----- 1106  
 DB 406 roGlyThrValGlnSerAlaAlaLeuMetAlaThrHisGlnProLeuSerLeuSerArgT 426  
 QY 1107 -----GCACCTGTACTG-----CCCC 1122  
 DB 426 hrThrAlaThrIleAlaProIleLeuSerValProLeuSerGlyGlnPheLeuGlnProA 446  
 QY 1123 CTGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1175  
 DB 446 laalaSerAlaAlaAlaAlaAlaAlaAlaAlaValGlnAlaLysTrpProAla 463

RESULT 10  
 ABR63546  
 ID ABR63546 standard; protein; 396 AA.  
 XX ABR63546;  
 AC ABR63546;  
 DT 18-SEP-2003 (first entry)  
 DE Danio rerio foxd3 homologue chicken CWH3.  
 DE Craniofacial malformation; variant foxd3; osteopathic; cardiant;  
 KW cytosstatic; auditory; neuroprotective; cleft palate syndrome;  
 KW congenital heart disease; cancer; chicken, CWH3.  
 OS Gallus gallus.  
 XX WO2003048196-A2.  
 XX 12-JUN-2003.  
 XX 09-DEC-2002; 2002WO-EP013936.  
 XX 07-DEC-2001; 2001EP-00129205.  
 XX (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEI.  
 XX Knapiak E, Sachdev S;  
 XX WPI; 2003-513737/48.  
 XX

PT New foxd3 polynucleotide and their (ant)agonists, useful for diagnosing  
 PT and treating craniofacial malformations, such as cleft palate syndrome,  
 PT loss of middle ear ossicles, congenital heart disease, cancer and mental  
 PT retardation.

XX Claim 34; Page 68-69; 70pp; English.

CC The present invention provides the protein and coding sequences of Danio  
 CC rerio foxd3. These can be used in the diagnosis and treatment of  
 CC craniofacial malformations, such as cleft palate syndrome, loss of middle  
 CC ear ossicles, congenital heart disease, Niikawa-Kuroki syndrome, cancer  
 CC and mental retardation. The present sequence is a homologue of the  
 CC zebrafish foxd3 protein, and is chicken CWH3

XX Sequence 396 AA;

Alignment Scores:  
 Pred. No.: 5,06e-20 Length: 396  
 Score: 445.00 Matches: 148  
 Percent Similarity: 41.41% Conservative: 40  
 Best Local Similarity: 32.60% Mismatches: 115  
 Query Match: 20.33% Indels: 152  
 DB: 6 Gaps: 21

US-10-087-080-31 (1-1209) x ABR63546 (1-396)

QY 49 GCGAGTACCTGGAGGCG-----GCGGGCGGAGGAGCGCGCTCCCGCTGTCG 99  
 DB 7 GlySerAspMetSerGlyGlnThrAlaLeuAlaGluAspValAspVal 26  
 QY 100 GCGCGGAGGAGCACTCCCTGGGTACATGGGAGTGGCGGCCAAG---CGTCCGCG 156  
 DB 27 GlyGluGlyAspAspAlaProGlyLysAspGlyAspGlyAlaArgSerProAlaAla 46  
 QY 157 -----GCGGCGCGCGCC 168  
 DB 47 LeuProArgLeuProLeuAspGluAlaAlaGluProGlyGluProGlyGlu 66  
 QY 169 AGAGATACGAGGCGGAGCGGAGCAAGAGTCCGGAGGCGG-----GCCGGCGCGAGGA 224  
 DB 67 SerGlySerGlySerGlySerProAlaProAlaGlyProGluGlyArgGlyGly 86  
 QY 225 GCGATCCGCGACAGCTGCTGACGGGTGGTGGCGGAGGCGCGGAGCGCGCGCGC 284  
 DB 87 Gly-----GlyGlyGlyGly-GlyGlyGluGluGlyAla 98  
 QY 285 GGGGCGAGCGCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 344  
 DB 98 rGlyGlyAlaAlaAlaAlaAlaAla-----AlaGlyGlnSerLysProLysSer 116  
 QY 345 GCGGCGGAGCGCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 404  
 DB 116 rLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaLeuGlnSerPr 136  
 QY 405 GGGCGGCGCTGACGCTGGGAGATCAACAGTACTCATGGCAAGTTCCTCTTTT 464  
 DB 136 oGlnLysLeuThrLeuSerGlyIleCysGluPheIleSerAsnArgPheProTyr 156  
 QY 465 CCGCGGAGCTACAGCGGCTGGCGCAACTCCGTGCGCCACCACTTCGCTCAACAGCTG 524  
 DB 156 rArgGluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCy 176  
 QY 525 CTTCGTACAGTGTGGCGACCCCTCGCGCTCGGCGGAGCGGAGCGGAGCGGAGCGG 584  
 DB 176 sPheValLysIleProArgGluProGlyAsnPro---GlyLysGlyAsnFyrIrpThrLe 195  
 QY 585 CAACCCCAACAGGAGTACACTTCGCGGAGCGGAGTTCGCGCGCGCGCGGAGCGGCT 644  
 DB 195 uAspProGlnSerGluAspMetPheAspAsnGlySerPheLeuArgArgArgLysArgPh 215  
 QY 645 CAGCCACCGCGCGCGG----- 660  
 DB 215 eLysArgHisGlnGlnGluHisLeuArgAspGlnThrAlaLeuMetMetGlnGlyPheG1 235

QY 661 -----GTCCCCCGCGCGCGCGCTGGCGGCGGAGGAGCGCGCGCGCGCGCG 710  
 DB 235 yAlaTyrGlyLeuAlaGlyProTyrGlyArgProTyr-----GlyLeuProProG1 252  
 QY 711 CCGCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCATCGCTCGCGCGCGCGCGCG 770  
 DB 252 yAlaTyrProHisProAlaAla----- 259  
 QY 771 GGAGCGCGCGCGCGCGCGCGCGCGCGAGTTCCTCAGTCTCTCGCCATCCACAGCATCTCGG 830  
 DB 259 -----ProValGlyProMetLeuProAlaValPro----- 277  
 QY 951 GGCCTGCTGCGCTCTCGCGGTACGGCGGCGGAGCGCGCGGCTGGCGCGCGCGCA 1010  
 DB 278 -----LeuLeuPro-----SerGlyGluLeuSerArg---LysAlaPheAs 290  
 QY 1011 GCGCGAGTGCACCGACCGCGCGCGCTCTCTCTTCACCTCTCCCGCGCGCGCGCG-- 1068  
 DB 290 nAlaGlnLeuGlyProSerLeu---GlnLeuGlnLeuSerSerLeuGlyAlaAlaGlySe 309  
 QY 1069 -----CCGCGCAAGCCA-----CTCCGAGG 1088  
 DB 309 rIleValLysSerGluProSerSerArgProSerPheSerIleGluAsnIleIleGlyG1 329  
 QY 1089 CCGCGCGCGCGCGCGCGCGCGCGCGCTGTACTGCGCC-----CTGCGGCTGCC 1133  
 DB 329 yProAlaAlaSerSerAla-----ProSerAlaGlnThrPheLeuArgProPr 345  
 QY 1134 GCGAGCTGCGAGCGCGCTTACTC---CGNCGTCTCT----- 1167  
 DB 345 oValThrValGlnSerGlyLeuValAlaHisGlnProLeuAlaLeuAlaArgThrThrAl 365  
 QY 1168 -----GGCCCGGACCTGTCTGTACCGCGGTGAGAGCGCTC 1200  
 DB 365 aAlaIleAlaProlleLeuSerValProThrAsnIleIle 378  
 RESULT 11  
 ABUS6559  
 ID ABUS6559 standard; protein; 409 AA.  
 XX AC ABUS6559;  
 XX DT 02-APR-2003 (first entry)  
 XX DE Lung cancer-associated polypeptide #152.  
 XX KW Lung cancer-associated polypeptide; cytostatic; emphysema;  
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.  
 XX OS Unidentified.  
 XX PN WO200286443-A2.  
 XX PD 31-OCT-2002.  
 XX PF 18-APR-2002; 2002WO-US012476.  
 XX PR 18-APR-2001; 2001US-0284770P.  
 PR 10-MAY-2001; 2001US-0290492P.  
 PR 09-NOV-2001; 2001US-0339245P.  
 PR 13-NOV-2001; 2001US-0350666P.

PR 29-NOV-2001; 2001US-0334370P.  
 PR 12-APR-2002; 2002US-0372246P.  
 XX (E0SB-) EOS BIOTECHNOLOGY INC.  
 PA  
 PI Aziz N, Murray R;  
 XX  
 DR WPI: 2003-093161/08.  
 XX N-PSDB; ABX76288.  
 XX  
 PT Detecting a lung cancer-associated transcript in a cell from a patient  
 PT for treating lung cancer, by contacting a biological sample from the  
 PT patient with a polynucleotide that exhibits increased or decreased  
 PT expression in lung cancer.  
 XX  
 PS Claim 27; Page 304; 453pp; English.  
 XX  
 CC The invention relates to a method for detecting a lung cancer-associated  
 CC transcript in a cell from a patient, comprising contacting a biological  
 CC sample from the patient with a polynucleotide that selectively hybridizes  
 CC to a sequence that is at least 80 % identical to a gene that exhibits  
 CC increased or decreased expression in lung cancer samples. Lung cancer-  
 CC associated polynucleotides and polypeptides are used for identifying a  
 CC compound that modulates a lung cancer-associated polypeptide, for  
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung  
 CC cancer in a patient and for treating a mammal having lung cancer by  
 CC administering a modulatory compound identified. The methods are useful  
 CC for treating lung cancer, such as small cell lung cancer, non-small cell  
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,  
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,  
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and  
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful  
 CC for diagnostic purposes and as targets for screening for therapeutic  
 CC compounds that modulate lung cancer, such as antibodies. Sequences  
 CC ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the  
 CC invention  
 XX  
 SX Sequence 409 AA;  
 XX

Alignment Scores:  
 Pred. No.: 1,73e-19 Length: 409  
 Score: 436.50 Matches: 145  
 Percent Similarity: 42.82% Conservative: 28  
 Best Local Similarity: 35.89% Mismatches: 124  
 Query Match: 19.94% Indels: 107  
 DB: 6 Gaps: 17

US-10-087-080-31 (1-1209) x ABUS6559 (1-409)

QY 196 AGTGGCGGAGGCGCGCGCGGAGGAGGCGATCCCGGACAGCTGCTGCACGGCTG 255  
 Db 35 ThrAlaGluSerGlyProProPro-----ProGlnProGluValLeuAlaThr 50  
 QY 256 GTGGCGGAGGCGCGGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 315  
 Db 51 ValLeuGluGluGlyGluThrAlaAlaGlyAlaGlyValProGlyGluAlaThrGly 70  
 QY 316 GAGGGGTGCA-----CGCAGCAGCATATACGCGCGCGCGCGCGCGCGCGCGCTACTCG 366  
 Db 71 ArgGlyAlaGlyGlyArgArgArgArgProLeuGlnArgGlyArgProProTrpSer 90  
 QY 367 TACATCGCGCTCATGCCATGCCATGCCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 426  
 Db 91 TyrIleAlaIleAlaMetAlaIleAlaHisAlaProGluArgArgLeuThrLeuGly 110  
 QY 427 GAGATCAACGAGTACTCATGGCAGTTCCTTTTCCGCGCGCGCGCTACACGGGCTGG 486  
 Db 111 GlyIleTyrIlePheIleThrGluArgPheProPheTyrArgAspAsnProLysIleTrp 130  
 QY 487 CGCACTCGGTGGCGCACACCTTTCGCTCAACGACTGCTTCTCAAGTGTGGCGGAC 546  
 Db 131 GlnAsnSerIleArgHisAsnLeuThrLeuAsnArgCysPheLeuIleProArgGlu 150

QY 547 CCCTCGCGCGCTGGGCGAAGGACAACTACTGATGCTCAACCCCAACGAGTACACC 606  
 Db 151 AlaGlyArgPro--GlyLysGlyAsnTyrTrpAlaLeuAspProAsnAlaGluAspMet 169  
 QY 607 TTGCGCGAGGGGCTCTCGCGCGCGCGCGCGCGCGCGCGCTCAGC----- 648  
 Db 170 PheGluSerGlySerPheLeuArgArgGlyArgPheLysArgSerLeuSerThr 189  
 QY 649 -----CACCGCGCGCGCGCTCCCGCGC----- 669  
 Db 190 TyrProAlaTyrMetHisAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 209  
 QY 670 -----CCGGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705  
 Db 210 AlaAlaAlaAlaAlaAlaIlePheProGlyAlaValProAlaAlaArgProProTyrPro 239  
 QY 706 -----GCCG 729  
 Db 230 GlyAlaValTyrAlaGlyTyrAlaProProSerLeuAlaAlaAlaAlaAlaAlaAla 249  
 QY 730 GCCCGCGCTCG 777  
 Db 250 ProAlaAlaSerPro-----GlyProCysArgValPheGlyLeuValProGluArg 266  
 QY 778 GCCAGCG 837  
 Db 267 ProLeuSerProGluLeuGlyProAlaProSerGlyProGlyGlySerCysAlaPheAl 286  
 QY 838 TTCGCGCA-----GCCGTCCCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCG 888  
 Db 286 AsrAlaGlyAlaProAlaThrThrThrGlyTyrGlnProAlaGly-----CysThrG 304  
 QY 889 GGCG---CCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 945  
 Db 304 YAlaArgPro-AlaAsnProSerAlaTyrAlaAlaAlaTyrAlaGlyProAspGlyAlaT 324  
 QY 946 TG---CAGGGCCCTGCTCGCGCTGCGCGCTACGCGCGCGCGCGCGCGCGCGCGCG 1002  
 Db 324 YrProGlnGlyAlaGlySerAlaIle-----PheAlaAlaAlaGlyA 338  
 QY 1003 GCGGCGGAGCGCGAGGTGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062  
 Db 338 rGleuAlaGly-----ProAlaSerProProAlaGlyGlySerSerGlyG 353  
 QY 1063 GC-----GGCG 1074  
 Db 353 lyValGluThrThrValAspPheTyrGlyArgThrSerProGlyGlnPheGlyAlaLeuG 373  
 QY 1075 AAGCACTCGAGCG 1134  
 Db 373 lyAlaCysTyrAsnProGlyGlyGln----- 381  
 QY 1135 GCAGCGCTCGAGCGCGCGCTTAGTCCGCGCGCTCTGG-----CCGCGACCTGTGTGTAC 1185  
 Db 382 -----LeuGlyGlyAlaSerAlaGlyAlaTyrHisAlaArgHisAlaAlaAlaTyrP 399  
 QY 1186 CCGGTGA 1193  
 Db 399 roGlyGly 401  
 RESULT 12  
 AAY43260  
 ID AAY43260 standard; protein; 553 AA.  
 XX  
 AC AAY43260;  
 XX  
 DT 17-JAN-2000 (first entry)  
 XX  
 DE Human FKHL7 protein sequence.  
 XX  
 KW FKHL7; human; forkhead transcription factor gene; diagnosis; therapy;  
 KW congenital heart disease.  
 XX





214	GC	CGCGGAGGAGCGATCCCGCGACGACGCTCTCCAGCGGTGTGTGGCGGAGCGCGGAG	273
	QY		
20	GlyGly	GlutInSerTyrTyrArgAlaAlaAlaAla-----AlaGlyGlyGlyTyr	37
	Db		
274	GC	CGCGGCGCGCGCGCGCGCGGCGG-----	297
	QY		
38	ThrAlaMetProAlaProMetSerValTyrSerHisProAlaHisAlaGluInTyrPro	57	
	Db		
298	GC	CGCGCGCGGAGCGCGGCGAGGCGTGCACGACGAAGCATATACGGCGGCGCGCCC-----	351
	QY		
58	GlyGlyMetAlaArgAlaTyrGly-----ProTyrThrProGlnProGlnPro	73	
	Db		
352	-----AAGCCCCCTACTCGPACATCGCGTCTATCGCCATCGCCATCGCGAC	399	
	QY		
74	LysAspMetValIysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleGlnAsn	93	
	Db		
400	TCGGCGGCGCGCGCTTACGCTGGGAGATCAACAGTACTCTCATGGCAAGTTCCTCC	459	
	QY		
94	AlaProAspLysIleThrLeuAsnGlyIleTyrGlnPheIleMetAspArgPhePro	113	
	Db		
460	TTTTTCCGGCGGACGTACACGGCGCTGGCGCAACTCCGTGGCGCACAACTTTTCGCTCAAC	519	
	QY		
114	PheTyrArgAspAsnIysGlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsn	133	
	Db		
520	GACTGCTTCGTCGAAGTGCTCGCGACCCCTCGCGGCGCTTGGGGCAAGGACAACTACTCG	579	
	QY		
134	GluCysPheValIysValProArgAspAspLysIysPro---GlyIysGlySerTyrTrp	152	
	Db		



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 29, 2004, 11:57:57 ; Search time 20.5 Seconds  
(without alignments)  
6089.346 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2189  
Sequence: 1 atgaagtggaggtgttcgt.....tsgagacgtctcctagttga 1209

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xl  
-Q=/cgn2\_1/USPTO\_spool/p/US10087080/runat\_29032004\_114527\_6585/app\_query.fasta\_1.1351  
-DB=Issued Patents AA -QFWT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pfo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087080@cgn 1 1 27 @runat\_29032004\_114527\_6585 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pcp:\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pcp:\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pcp:\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pcp:\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS.COMB.pcp:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pcp:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	431.5	19.7	553	US-09-083-351-2	Sequence 2, Appli
2	431.5	19.7	553	US-09-083-352-2	Sequence 2, Appli
3	394.5	18.0	330	US-09-976-594-928	Sequence 928, App
4	387	17.7	473	US-08-857-076-99	Sequence 99, Appl
5	347	15.9	480	US-07-882-292-2	Sequence 2, Appli
6	347	15.9	480	US-08-331-644-2	Sequence 2, Appli
7	347	15.9	480	PCT-US93-04102-2	Sequence 100, App
8	336	15.3	347	US-08-857-076-100	Sequence 14, Appl
9	333	15.2	106	US-09-083-351-14	Sequence 14, Appl
10	333	15.2	106	US-09-083-352-14	Sequence 15, Appl
11	332	15.2	106	US-09-083-351-15	Sequence 15, Appl
12	332	15.2	106	US-09-083-352-15	Sequence 15, Appl

13	329	15.0	544	4	US-09-087-134-14	Sequence 14, Appl
14	320.5	14.6	106	3	US-09-083-351-12	Sequence 12, Appl
15	320.5	14.6	106	3	US-09-083-352-12	Sequence 12, Appl
16	320.5	14.6	2294	4	US-09-252-991A-17231	Sequence 17231, A
17	319	14.6	365	3	US-09-113-309-2	Sequence 2, Appli
18	319	14.6	365	3	US-09-521-109-2	Sequence 2, Appli
19	319	14.6	365	4	US-09-562-332-2	Sequence 2, Appli
20	317	14.5	369	4	US-09-252-991A-25394	Sequence 25394, A
21	315	14.4	663	4	US-09-252-991A-30843	Sequence 30843, A
22	314.5	14.4	106	3	US-09-083-351-13	Sequence 13, Appl
23	314.5	14.4	106	3	US-09-083-352-13	Sequence 13, Appl
24	312.5	14.3	783	4	US-09-252-991A-18035	Sequence 18035, A
25	311.5	14.2	866	4	US-09-252-991A-20509	Sequence 20509, A
26	311	14.2	863	4	US-09-252-991A-26099	Sequence 26099, A
27	307	14.0	838	4	US-09-252-991A-27068	Sequence 27068, A
28	307	14.0	957	4	US-09-252-991A-20408	Sequence 20408, A
29	307	14.0	1706	4	US-09-252-991A-31760	Sequence 31760, A
30	305.5	14.0	106	3	US-09-083-351-7	Sequence 7, Appli
31	305.5	14.0	106	3	US-09-083-352-7	Sequence 7, Appli
32	305.5	14.0	534	4	US-09-087-134-11	Sequence 11, Appl
33	304.5	13.6	446	4	US-09-252-991A-27110	Sequence 27110, A
34	304.5	13.9	518	3	US-09-113-309-19	Sequence 19, Appl
35	304.5	13.9	518	4	US-09-562-332-19	Sequence 19, Appl
36	304.5	13.9	518	4	US-09-562-332-19	Sequence 19, Appl
37	303	13.8	467	4	US-09-252-991A-18296	Sequence 18296, A
38	303	13.8	586	4	US-09-252-991A-24514	Sequence 24514, A
39	303	13.8	720	4	US-09-252-991A-21881	Sequence 21881, A
40	302.5	13.8	726	4	US-09-252-991A-20675	Sequence 20675, A
41	302	13.8	904	4	US-09-976-594-615	Sequence 615, App
42	301.5	13.4	595	4	US-09-370-838-187	Sequence 187, App
43	298	13.6	809	4	US-09-252-991A-31759	Sequence 31759, A
44	297.5	13.2	467	4	US-09-252-991A-18296	Sequence 18296, A
45	296.5	13.5	495	4	US-09-252-991A-31949	Sequence 31949, A

#### ALIGNMENTS

RESULT 1  
US-09-083-351-2  
; Sequence 2, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,351  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UIA-029.02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000

272 rProProGlySerLeuProSerAlaArgProLeuSerLeuAspGlyAlaAspSerAlaPr 292

895 GGGCCCTGCCCGCGTCCCAGGTTCCTCCCGGCCTCCTCCC----- 935

292 oProPro-ProAla-----ProSerAlaProProProHisHisSerGlnGlyP 308

936 -----CGCGCGCCCTCGACGGCCCTCGACGGCCCTCGCTGCCGCTC 966

308 heSerValAspAanIleMetThrSerLeuArgGlySerProGlnSerAlaAlaGluL 328

967 TGGCGGTACGGCGCGGCGAGCGCGCGGTGGCGCGCGCGAGCGCGAGGTGCCACCG 1026

328 eu-----SerSerGlyLeuLeuAlaSerAlaAlaSerS 340

1027 ACGCGGCC-----GCCCTCCTC----- 1043

340 exArgAlaGlyIleAlaProProLeuAlaLeuGlyAlaTySerProGlyGlnSerSerL 360

1044 -----GCTGTCACTCTCCCGCGCGCGCGCCCCCGCC 1074

360 euTySerSerProCysSerGlnThrSerSerAlaGlySerSerGlyGlyGlyGlyG 380

1075 AAGCCACTCGAGCGCGCGCGCGCGCGCACCTGTAC---TGCCCCCTGCCG--- 1128

380 lyAlaGlyAlaalaGlyGly-AlaGlyGlyAlaGlyThrTyHisCysAsnLeuGlnAla 399

1129 -----CTGCCCCGCGAGCCCTG 1143

400 MetSerLeuTyAlaAlaGlyGluArgGlyGlyHisLeuGlnGlyAlaProGlyGlyAla 419

1144 CAGCGCGCTTAGTCGCGCTGCTGCGCGCGCACCTGTGTACCCTGGTGGAGACG 1197

420 GlyGlySerAlaValAspAsnProLeuProCaspTySerLeuProvalThr 437

RESULT 2  
US-09-083-352-2  
; Sequence 2, Application US/09083352  
; Patent No. 6207450  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patel, Shiva  
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS  
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,352  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UIA-029,01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 amino acids

QY	285	GGGCGACGCGGGGGCGGGGGAGCGGGTGCACGCAAGCAGCATATACGCGG	345
Db	8	GlyProGlyAla-----	15
		-----ProThrThrGlu	
QY	345	CGGCGCAAGCCCCCTACTCTGATCATCGCCTCATCGCCATGCCATCGGACATCGGGG	405
Db	16	ProThrLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleGlnSerSerPro	35
QY	406	GGCGGGGCGCTTGAGCTGGGGGAGATCAACAGCATCTCATGGCGAAGTTCGCCCTTTTC	465

```

Db      36 GlyGlnArgAlaThrLeuSerGlyIleTyrArgTyrIleMetGlyArgPheAlaPheTyr 55
QY      466 CGCGGAGCTACACGGCTGGCGCACTCCGCGCCACAACTTCGCTCAACGACTGC 525
Db      56 ArgHisAsnArgProGlyTyrGlnHisSerIleArgHisAsnLeuSerLeuAsnGluCys 75
QY      526 TTCGTAAGTCTGCGGACCCCTCGCGGCCCTGGGGCAAGACAACTACTGATGCTC 585
Db      76 PheValLysValProArgAspArgLysPro---GlyLysGlySerTyrTrpThrLeu 94
QY      586 AACCCCAACAGCAGGTACACCTTCGCCGACGGGGTCTTCGCCCGCGCGCGCAAGCGCTC 645
Db      95 AspProAspCysHisAspMetPheGluHisGlySerPheLeuArgArgArgArgPhe 114
QY      646 AGCCAC-----CGCGCGCGTCCCGCG-----CCCGGG 675
Db      115 ThrArgGlnThrGlyAlaGluGlyThrArgGlyProAlaLysAlaArgArgGlyProLeu 134
QY      676 CTGCGGCGCGAGAGCGCCCGCGCTCCCGCGCGCG-----711
Db      135 ArgAlaThrSerGlnAspProGlyValProAsnAlaThrThrGlyArgGlnCysSerPhe 154
QY      712 CGCGCGCGCGCGCGCG-----GCGCGCGCG 735
Db      155 ProProGluLeuProAspProLysGlyLeuSerPheGlyLeuValGlyAlaMetPro 174
QY      736 GCCTCG-----CCCGCATGCGCTCGCGCGCGCGCGCGAGAG 774
Db      175 AlaSerMetCysProAlaThrThrAspGlyArgProArgProMetGluProLysGlu 194
QY      775 CGCGCGCGCGCGCGAGTCTCCAGCTCTTCGCCCATCGACATCCTCGCGCAAG 834
Db      195 IleSerThrPro-----Lys 199
QY      835 CCTTTCGCGAGCGCTCGCCTCAGGACACGCGCGCGCGCG---ACGAGCTTTCAGTGGCG 891
Db      200 Pro-----AlaCysProGlyGluLeuProValAlaThrSer 211
QY      892 GCGCGCGCGCGCGCGCGCTCGCGCGTTCGCGCG-----CTCCTC 933
Db      212 SerSerSerCysProAlaPhe---GlyPheProAlaGlyPheSerGluAlaGluSerPhe 230
QY      934 CGCGCGCGCGCGCTGACGGCGCGCTGCGCGCTCTGCGCTACGCGCGCGCGCGCGCGCG 993
Db      231 AsnLysAlaProThrProValLeuSerProGluSerGlyIleGlySerSerTyrGlnCys 250
QY      994 CGGCTG-----999
Db      251 ArgLeuGlnAlaLeuAsnPheCysMetGlyAlaAspProGlyLeuGluHisLeuLeuAla 270
QY      1000 GCGCGCGCGAGCGAGGTCCACCGACCGCGCGCGCGCGCTCTGCTTGCACCTCTCCCG 1059
Db      271 SerAlaAlaProSerProAlaProProThrProProGlySerLeuArgAlaProLeuPro 290
QY      1060 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
Db      291 -----LeuProThrAspHisLysGluProTrpValAlaGlyPheProValGlnGly 308
QY      1120 CCCTGCGCGCTGCGCGAGCGCTTCAGCGCGCGCTTAGTCCGCGCTCTCGC 1170
Db      309 GlySerGlyTyrProLeuGlyLeuThrProCysLeuTyrArgThrProGly 325

```

## RESULT 4

```

US-08-857-076-99
; Sequence 99, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne

```

```

; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/957,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-857-076-99

```

## Alignment Scores:

```

Pred. No.: 5,77e-16 Length: 473
Score: 387.00 Matches: 124
Percent Similarity: 44.36% Conservative: 45
Best Local Similarity: 32.55% Mismatches: 132
Query Match: 17.68% Indels: 80
DB: 3 Gaps: 15

```

US-10-087-080-31 (1-1209) x US-08-857-076-99 (1-473)

```

QY      70 GCGCGCAGCGAGCGCGCTCCCGCTGTCGCGCGCGCGGACGACTCCCTGGGCTCAGAT 129
Db      88 GlySerAlaGlyAlaMetAsnSerMetThrAlaAlaGlyValThrAlaMetGlyThr--- 106
QY      130 GGGGACTGCGCGCGCAAGCGCTCGCGCGCGCGCGCGCGCGCGAGATACGCGGCGCGCG 189
Db      107 -----AlaLeuSerProSer---GlyMetGlyAlaMetGlyAlaGlnAlaAla 222
QY      190 GAACAGAGTGGCGGAGCGCGCGCGCGCGAGGAGCGATCCCGCGAGCAGCTGTGTGCA 249
Db      123 SerMetMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCysMetSerProMet 142
QY      250 GCGGTGTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309
Db      143 AlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGly-----GlyGlyGlyAsp 159
QY      310 AGCGCGAGGAGGTGCACGACGACCAAGCCATATACGCGCGCGCGCGCGCGCGCGCGCG 369
Db      160 AlaLysThrPheLysArgSerTyrProHisAla-----LysProProTyrSerTyr 176
QY      370 ATCCGCTCATGCCATGCGCATCCGCGACTCCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 429
Db      177 IleSerLeuIleThrMetAlaIleGlnArgAlaProSerLysMetLeuThrLeuSerGlu 196
QY      430 ATCAACGAGTACTCATGGCAAGTTCCCTTTTCGCGCGCGAGCTACACGGGCTGGCGC 489
Db      197 IleTyrGlnTrpIleMetAspLeuPheProTyrTyrArgGlnAsnGlnArgTrpGln 216
QY      490 AACTCCGTCGCGCACAACTTCGCTCAACGACTGCTTGTCAAGTGTCTGCGCGCGCGCGCG 549
Db      217 AsnSerIleArgHisSerLeuSerPheAsnAspCysPheValLysValAlaArgSerPro 236
QY      550 TCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
Db      237 AspLysPro---GlyLysGlySerTyrTrpThrLeuHisProAspSerGlyAsnMetPhe 255
QY      610 GCGGACGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
Db      256 GluAsnGlyCysTyrLeuArgArgGlnLysArg-----266
QY      670 CCCGCGCTGCGCGCGCGAGGAGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCG 729
Db      267 -----PheLysCysGluLysGlnProGly-----AlaGlyGly 277
QY      730 GCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783
Db      278 GlyGlySerGlySerGlySerGlyAlaLysGlyGlyGlyProGluSerArgLysasp 297

```

```

QY 784 CCGCGGGCAAGTCTCCAGCTCTTCCGATCGACAGCATCTGGCGAAGCCCTTCGCG 843
Db 784 CCGCGGGCAAGTCTCCAGCTCTTCCGATCGACAGCATCTGGCGAAGCCCTTCGCG 843
QY 298 ProSerGly-----AlaSerAsnProSerAlaAspSerProLeuHisArgGlyValHis 315
Db 298 ProSerGly-----AlaSerAsnProSerAlaAspSerProLeuHisArgGlyValHis 315
QY 844 AGCGTCCGCTCAGGACAGCGCCCGCGGACAGACGCTTCAGTGGGGCGCGCCCTGCG 903
Db 844 AGCGTCCGCTCAGGACAGCGCCCGCGGACAGACGCTTCAGTGGGGCGCGCCCTGCG 903
QY 316 -----GlyLeuThrGlyGlnLeuGluGlyAlaProAla 326
Db 316 -----GlyLeuThrGlyGlnLeuGluGlyAlaProAla 326
QY 904 CCGCGGCTGCGCGGTTCCCGGCTCTCCCGCGCGCGCCCTGCGAGGCGCTGCTGCGG 963
Db 904 CCGCGGCTGCGCGGTTCCCGGCTCTCCCGCGCGCGCCCTGCGAGGCGCTGCTGCGG 963
QY 327 ProGlyProAlaAlaSerProGlnThrLeuAsp----- 337
Db 327 ProGlyProAlaAlaSerProGlnThrLeuAsp----- 337
QY 964 CTCTGCGGTACGGCGGCGGAGCGCGCGGCTGGCGCGCGGCGGCGGAGCGCGAGGTGCA 1023
Db 964 CTCTGCGGTACGGCGGCGGAGCGCGCGGCTGGCGCGCGGCGGCGGAGCGCGAGGTGCA 1023
QY 338 -----HisSerGlyAlaThrAlaThrGlyGlyAlaSerGlnLeuLysThrPro 353
Db 338 -----HisSerGlyAlaThrAlaThrGlyGlyAlaSerGlnLeuLysThrPro 353
QY 1024 -----CCGACCGCGCGCGCCCTC-----CTGCTTGCACCTCTCCCGCGG 1062
Db 1024 -----CCGACCGCGCGCGCCCTC-----CTGCTTGCACCTCTCCCGCGG 1062
QY 354 AlaSerThrAlaProProLieserSerGlyProGlyAlaLeuAlaSerValProAla 373
Db 354 AlaSerThrAlaProProLieserSerGlyProGlyAlaLeuAlaSerValProAla 373
QY 1063 GCGGCGCGCGCGGAGCGCTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCTGTACTGCCCC 1122
Db 1063 GCGGCGCGCGCGGAGCGCTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCTGTACTGCCCC 1122
QY 374 SerHisProAlaHisGlyLeu---AlaProHisGluSerGlnLeuHisLeu----- 389
Db 374 SerHisProAlaHisGlyLeu---AlaProHisGluSerGlnLeuHisLeu----- 389
QY 1123 CTGCGGCTGCGCGCGCGCTGCGAGCGCGCTTAGTCGCGCGCTCTGCGCGCGCGCGCG 1182
Db 1123 CTGCGGCTGCGCGCGCGCTGCGAGCGCGCTTAGTCGCGCGCTCTGCGCGCGCGCGCG 1182
QY 390 -----LysGlyAspProHisTyrSer 396
Db 390 -----LysGlyAspProHisTyrSer 396
QY 1183 TAC 1185
Db 1183 TAC 1185
QY 397 Phe 397
Db 397 Phe 397

```

## RESULT 5

```

US-07-882-292-2
; Sequence 2, Application US/07882292
; Patent No. 5324638
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: c/o Cooper and Dunham, 30 Rockefeller
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/882,292
; FILING DATE: 19920513
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-977-9550
; TELEFAX: 212-664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

```

## MOLECULE TYPE: protein

US-07-882-292-2

## Alignment Scores:

```

Pred. No.: 1,41e-13 Length: 480
Score: 347.00 Matches: 124
Percent Similarity: 40.24% Conservative: 41
Best Local Similarity: 30.24% Mismatches: 144
Query Match: 15.85% Indels: 101
DB: 1 Gaps: 17

```

US-10-087-080-31 (1-1209) x US-07-882-292-2 (1-480)

```

QY 22 CTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
Db 22 CTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
QY 81 ProProGlnAlaArgGlyAlaProAlaAlaAspAspLysGly-----CACTCC 117
Db 81 ProProGlnAlaArgGlyAlaProAlaAlaAspAspLysGly-----CACTCC 117
QY 96 ---ProGlnProLeuLeuLeuProProSerAlaAlaLeuAspGlyAlaLysAlaAspAla 114
Db 96 ---ProGlnProLeuLeuLeuProProSerAlaAlaLeuAspGlyAlaLysAlaAspAla 114
QY 118 CTGGCTCAGATGGGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 177
Db 118 CTGGCTCAGATGGGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 177
QY 115 LeuGlyAlaLysGlyGlu-----ProGlyGlyGlyProAlaGluLeuAlaPro 130
Db 115 LeuGlyAlaLysGlyGlu-----ProGlyGlyGlyProAlaGluLeuAlaPro 130
QY 178 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 237
Db 178 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 237
QY 131 ValGlyProAspGlnLysGlyAlaGlyAlaGlyGlyGlu----- 146
Db 131 ValGlyProAspGlnLysGlyAlaGlyAlaGlyGlyGlu----- 146
QY 238 GCAGCTGTGACGCGGTGTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
Db 238 GCAGCTGTGACGCGGTGTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
QY 147 -----LysLysGlyAlaGlyGlyLys 155
Db 147 -----LysLysGlyAlaGlyGlyLys 155
QY 298 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357
Db 298 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357
QY 156 AspGlyGluGlyGlyGlyGlyGly-----AspLysAsnAsnGlyLysTyrGluLysPro 173
Db 156 AspGlyGluGlyGlyGlyGlyGly-----AspLysAsnAsnGlyLysTyrGluLysPro 173
QY 358 CCTACTCTGATCGCGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
Db 358 CCTACTCTGATCGCGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
QY 174 ProPheThrTyrAsnAlaLeuLeuMetMetAlaLeuArgGlnSerProGluLysArgLeu 193
Db 174 ProPheThrTyrAsnAlaLeuLeuMetMetAlaLeuArgGlnSerProGluLysArgLeu 193
QY 418 AGCTGGCGGAGATCAACAGTACCTCATGGCAAGTTCCCTTTTCCCGCGCGCGCGCG 477
Db 418 AGCTGGCGGAGATCAACAGTACCTCATGGCAAGTTCCCTTTTCCCGCGCGCGCGCG 477
QY 194 ThrLeuAsnGlyLeuTyrGluPheLeuMetLysAsnPheProTyrTyrArgGluAsnLys 213
Db 194 ThrLeuAsnGlyLeuTyrGluPheLeuMetLysAsnPheProTyrTyrArgGluAsnLys 213
QY 478 AGCGCTGGCGCAACTCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537
Db 478 AGCGCTGGCGCAACTCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537
QY 214 GlnGlyTyrGlnAsnSerIleArgHisAsnLeuSerLeuAsnLysCysPheValLysVal 233
Db 214 GlnGlyTyrGlnAsnSerIleArgHisAsnLeuSerLeuAsnLysCysPheValLysVal 233
QY 538 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597
Db 538 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597
QY 234 ProArgHisTyrAspAspPro---GlyLysGlyAsnTyrTrpMetLeuAspProSerSer 252
Db 234 ProArgHisTyrAspAspPro---GlyLysGlyAsnTyrTrpMetLeuAspProSerSer 252
QY 598 GAGTACACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
Db 598 GAGTACACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
QY 253 AspAspValPheLeuGlyGlyThrThrGlyLysLeuArgArgArgSerThrThrSerArg 272
Db 253 AspAspValPheLeuGlyGlyThrThrGlyLysLeuArgArgArgSerThrThrSerArg 272
QY 643 -----CTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690
Db 643 -----CTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690
QY 273 AlaLysLeuAlaPheLeuArgAlaArgLeuThrSerThrGlyLeuThrPheMetAsp 292
Db 273 AlaLysLeuAlaPheLeuArgAlaArgLeuThrSerThrGlyLeuThrPheMetAsp 292
QY 691 GCCCGCGCGCG-----CTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 723
Db 691 GCCCGCGCGCG-----CTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 723
QY 293 ArgAlaGlySerLeuTyrTrpProMetSerProPheLeuSerLeuHisHisProArgAla 312
Db 293 ArgAlaGlySerLeuTyrTrpProMetSerProPheLeuSerLeuHisHisProArgAla 312
QY 724 -----CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Db 724 -----CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
QY 313 SerSerThrLeuSerTyrAsnGlyThrThrSerAlaTyrProSerHisProMetProTyr 332
Db 313 SerSerThrLeuSerTyrAsnGlyThrThrSerAlaTyrProSerHisProMetProTyr 332
QY 754 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
Db 754 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
QY 333 SerSerValLeuThrGlnAsnSerLeuGlyAsnAsnHisSerPheSerThrAlaAsnGly 352
Db 333 SerSerValLeuThrGlnAsnSerLeuGlyAsnAsnHisSerPheSerThrAlaAsnGly 352

```



QY 907 -----CGCGTGC-----CGCGTTC 921  
 Db 393 LeuAsnPro-CysSerValAsnLeuLeuAlaGlyGlnThrSerTyrPheProHisVa 412  
 QY 922 CCGCGCTCTCTCCCGCGCGCCCTGCAGGCGCCCTGCTGCGCTCTGCGGTACGCGCG 981  
 Db 412 1ProHisProSer-----MetThrSerG1 420  
 QY 982 GCGAGCGCGCGCGTGGCGCGCGAGCGAGCGGAGTGCACCGACCGCGCGCGCGCTC 1041  
 Db 420 nThrSerThrSerMetSerAlaArgAlaAlaSerSerThrSerProGlnAlaProSe 440  
 QY 1042 CTGCTTGACCTCTCCCGCGCGCGCC 1069  
 Db 440 rThrLeuProCysGluSerLeuArgPro 449

## RESULT 7

PCT-US93-04102-2  
 ; Sequence 2, Application PC/TUS9304102  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tao, Wufan  
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
 ; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John P. White  
 ; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/04102  
 ; FILING DATE: 19930430  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/882,292  
 ; FILING DATE: 13-MAY-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 41472A-PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-977-9550  
 ; TELEFAX: 212-664-0525  
 ; TELEX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 480 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 PCT-US93-04102-2

Alignment Scores:  
 Pred. No.: 1,41e-13 Length: 480  
 Score: 347.00 Matches: 124  
 Percent Similarity: 40.24% Conservative: 41  
 Best Local Similarity: 30.24% Mismatches: 144  
 Query Match: 15.85% Indels: 101  
 DB: 5 Gaps: 17

US-10-087-080-31 (1-1209) x PCT-US93-04102-2 (1-480)

QY 22 CTCGCGCGCGCCACCGGACAGCAGCGGCGTGCCTGGAGGCGCGCGCGCGCGAC 81  
 Db 81 ProGlnAlaArgGlyAlaProAlaAlaAspAspLysGly-----MetThrSerG1 420

QY 82 GCGCGCTCCCGCTG-----TCGGCGCGCGGAGAC-----GACTCC 117  
 Db 96 ---ProGlnProLeuLeuLeuProSerAlaAlaLeuAspGlyAlaLysAlaAspAla 114  
 QY 118 CTGGCTCAGATGGGACTGCGCGGCAAGCCTCGCGGGCGGCGCGCGGAGATACG 177  
 Db 115 LeuGlyAlaLysGlyGlu-----ProGlyGlyGlyProAlaGluLeuAlaPro 130  
 QY 178 CAGGCGACGCGCAACAGAGTGGCGGAGGCGCGCGCGCGGAGGAGGCGATCCCGGCA 237  
 Db 131 ValGlyProAspGlnLysGluLysGlyAlaGlyAlaGlyGluGlu----- 146  
 QY 238 GCAGCTCTGCAGCGGTGTGGCGGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCG 297  
 Db 147 -----LysLysGlyAlaGlyGlyGlyLys 155  
 QY 298 GCGCGCGCGGAGCGCGGAGGTGCACGACGACCATATACGGCGCGCGCCCAAGCCC 357  
 Db 156 AspGlyGluGlyLysGluGly-----AspLysAsnAsnGlyLysTyrGluLysPro 173  
 QY 358 CCTTACTCGTACATCGCGCTCATCGCATCGCATCGCGACTCGCGCGCGCGCGGTTG 417  
 Db 174 ProPheThrTyrAsnAlaLeuMetMetAlaIleArgGlnSerProGluLysArgLeu 193  
 QY 418 AGCTGCGGAGATCAACGAGTACCTCATGGCAAGTCCCTTTTCCGCGCGCGCTAC 477  
 Db 194 ThrLeuAsnGlyIleTyrGluPheIleMetLysAsnPheProTyrTyrArgGluAsnLys 213  
 QY 478 ACGGCTGGCGCAACTCCGTCGCGCCACAACTTCGCTCAACGACTCTTCGTCAGG 537  
 Db 214 GlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnLysCysPheValLysVal 233  
 QY 538 CTGCGCGACCCCTCGCGCGCTGGGCAAGGACAACTACTGGATGCTCAACCCCAAGC 597  
 Db 234 ProArgHisTyrAspAspPro---GlyLysGlyAsnTyrTrpMetLeuAspProSer 252  
 QY 598 GAGTACACCTTCGCGCGCGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642  
 Db 253 AspAspValPheIleGlyGlyThrThrGlyLysLeuArgArgSerThrThrSerArg 272  
 QY 643 -----CTCAGCCACCGCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 690  
 Db 273 AlaLysLeuAlaPheLysArgAlaArgLeuThrSerThrGlyLeuThrPheMetAsp 292  
 QY 691 GCCCGCGCG-----CTCCCGCGCGCGCGCG-----CGCGCGCGCG 723  
 Db 293 ArgAlaGlySerLeuTyrTrpProMetSerProPheLeuSerLeuHisHisProArgAla 312  
 QY 724 -----CCCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753  
 Db 313 SerSerThrLeuSerTyrAsnGlyThrThrSerAlaTyrProSerHisProMetProTyr 332  
 QY 754 TCGCGCGCGCGCGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807  
 Db 333 SerSerValLeuThrGlnAsnSerLeuGlyAsnHisSerPheSerThrAlaAsnGly 352  
 QY 808 TTCGCCATCAGACATCTCGCGCAAG-----CCCTTCGCGCGCGCGCTCGCGCGCG 861  
 Db 353 LeuSerValAspArgLeuValAsnGlyGluIleProTyrAlaThrHisHisLeuThrAla 372  
 QY 862 ACGCGCGCGCGCGCGCTTCAGTGGCG-----CGCGCGCGCGCGCGCG----- 906  
 Db 373 AlaAlaLeuAlaAlaSerValProCysGlyLeuSerValProCysSerGlyThrTyrSer 392  
 QY 907 -----CGCGTGC-----CGCGGTTTC 921  
 Db 393 LeuAsnPro-CysSerValAsnLeuLeuAlaGlyGlnThrSerTyrPhePheProHisVa 412  
 QY 922 CCGCGCTCTCTCCCG 981  
 Db 412 1ProHisProSer-----MetThrSerG1 420

QY 982 GCGAGCGCGCGGCTGCGCGCGGAGCGGAGGTCACCGAGCGCGCGCGCCCTC 1041  
Db 420 nThrsrThrSerMetSerAlaArgAlaAlaSerSerThrSerProGlnAlaProSe 440  
QY 1042 CTGCTTGACACCTCTCCCGCGCGCGCGCC 1069  
Db 440 rThrsrProCysGluSerLeuArgPro 449

RESULT 8  
US-08-857-076-100  
; Sequence 100, Application US/08857076C  
; Patent No. 6225120  
; GENERAL INFORMATION:  
; APPLICANT: Kimura, Koutarou  
; APPLICANT: Patterson, Garth  
; APPLICANT: Osg, Scott  
; APPLICANT: Paradis, Suzanne  
; APPLICANT: Tissenbaum, Heidi  
; APPLICANT: Morris, Jason  
; APPLICANT: Kowsek, Allison  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
; IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
; FILE REFERENCE: 00786/351001  
; CURRENT APPLICATION NUMBER: US/08/857,076C  
; CURRENT FILING DATE: 1997-05-15  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 100  
; LENGTH: 347  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-857-076-100

Alignment Scores:  
Pred. No.: 6,31e-13 Length: 347  
Score: 336.00 Matches: 87  
Percent Similarity: 47.35% Conservative: 29  
Best Local Similarity: 35.51% Mismatches: 102  
Query Match: 15.35% Indels: 27  
DB: 3 Gaps: 7

US-10-087-080-31 (1-1209) x US-08-857-076-100 (1-347)  
QY 208 GCGCGCGCGCGGAGGAGGCGATCCCGGAGCAGCTGCTGCAGCGTGTGCGGAGGCG 267  
Db 62 GlyProLeuAlaProAlaProAlaProAlaProLeuGlyProThrPheProGlyLeuGly 81  
QY 268 GCGAGCG 321  
Db 82 LeuSerGlyGlySerSerSerSerGlyTyrglyAlaProGlyProGlyLeuValHisGly 101  
QY 322 GCAGCGAGCAGCATATACCG 372  
Db 102 LysGluMetProLysGlyTyra-GAlaProAlaHisAlaLysProProTyrsrTyrlle 121  
QY 373 GCGCTCATCCGATGCGCATCCCGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432  
Db 122 SerLeuIleThrMetAlaIleGlnAlaProGlyLysValLeuThrLeuSerGluIle 141  
QY 433 AACGAGTACTCATGGCAAGTTCCTTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492  
Db 142 TyrGlnTrpIleMetAspLeuPheProTyraArgAspAsnGlnGlnArgTrpGlnAen 161  
QY 493 TCGTGCGGCACACCTTTTCGCTCAACGACTGCTTCTCAAGGTGCTGCGCGCGCGCGCGCG 552  
Db 162 SerIleArgHisSerLeuSerPheAsnAspCysPheValLysValAlaArgSerProAsp 181  
QY 553 CGGCGCTGGGCGAGGACAACTACTGATGCTCAACCCCAACAGCGAGTACACCTTCGCG 612  
Db 182 LysPro--GlyLysGlySerTyTrpAlaLeuHisProSerSerGlyAsnMetPheGlu 200  
QY 613 GAGCGGTCTTCG 672

Db 201 AsnGlyCysTyrrLeuArgArgGlnLysArgPheLysLeuGluLysVal-LysLysGI 220  
QY 673 GCGCTGCGCGCGCGGAGGCG 708  
Db 220 yGlySerGlyAlaSerThrThrArgaenGlyThrGlySerAlaAlaSerThrThrPr 240  
QY 709 GCG 768  
Db 240 aAlaAlaThrValThrSerProGlnProPro----- 252  
QY 769 GAGGAGCG 828  
Db 253 ----ProAlaProGluProGluAlaGlnGly---GlyGluAspValGlyAlaLeuaspCy 270  
QY 829 CGCAAGCCCTTCGCGAGCG 882  
Db 270 sGlySerProAlaSerSerThrProTyrrPheThrGlyLeuGluLeuProGlyAspLeuLy 290  
QY 883 CAGTGGCG 895  
Db 290 sLeuAspAlaPro 294

RESULT 9  
US-09-083-351-14  
; Sequence 14, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TRANSCRIPTION FACTOR  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELLIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,351  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UTA-029.02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 106 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-083-351-14  
Alignment Scores:  
Pred. No.: 9.13e-13 Length: 106  
Score: 333.00 Matches: 53  
Percent Similarity: 79.00% Conservative: 16

```
Best Local Similarity: 63.00% Mismatches: 19
Query Match: 15.21% Indels: 2
DB: 3 Gaps: 2
US-10-087-080-31 (1-1209) x US-09-083-352-14 (1-106)
QY 343 CGCGGCCCC---AAGCCCCCTACTCGTACATCGGCTCATCGCCATCGCCATCGCCGAC 399
Db 2 ArgArgProGluLysProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSer 21
QY 400 TCGGGCGGGCGGCTTACAGCTGGCGGAGATCAACGAGTACCTCATCGGCAAGTTCCCC 459
Db 22 SerProThrLysArgLeuThrLysSerGluLeuTyrGlnPheLeuGlnSerArgPhePro 41
QY 460 TTTTTCGGCGGAGTACACAGGCTGGCGCACTCCGTCGCCCAACACTTTCCTCAAC 519
Db 42 PhePheArgGlySerTyrGlnGlyTrpLysAsnSerValArgHisAsnLeuSerLeuAsn 61
QY 520 GACTGCTTCGTCAAGTCTGCGGACCCCTCGGCGCCCTGGGCAAGCAACTACTGG 579
Db 62 GlucysPheIleLysLeuProLysGlyLeuGlyArgPro---GlyLysGlyHisTyrTrp 80
QY 580 ATGCTCAACCCCAACAGCGAGTACACTTCGCGCGGGGTCTTCGCGCCGCCCGCAAG 639
Db 81 ThrIleAspProAlaSerGluPheMetPheGluAsnGlySerPheArgArgArgArg 100
RESULT 10
US-09-083-352-14
; Sequence 14, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-083-352-14
Alignment Scores: 9.13e-13 Length: 106
Pred. No.:

Score: 333.00 Matches: 63
Percent Similarity: 79.00% Conservative: 16
Best Local Similarity: 63.00% Mismatches: 19
Query Match: 15.21% Indels: 2
DB: 3 Gaps: 2
US-10-087-080-31 (1-1209) x US-09-083-352-14 (1-106)
QY 343 CGCGGCCCC---AAGCCCCCTACTCGTACATCGGCTCATCGCCATCGCCATCGCCGAC 399
Db 2 ArgArgProGluLysProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSer 21
QY 400 TCGGGCGGGCGGCTTACAGCTGGCGGAGATCAACGAGTACCTCATCGGCAAGTTCCCC 459
Db 22 SerProThrLysArgLeuThrLysSerGluLeuTyrGlnPheLeuGlnSerArgPhePro 41
QY 460 TTTTTCGGCGGAGTACACAGGCTGGCGCACTCCGTCGCCCAACACTTTCCTCAAC 519
Db 42 PhePheArgGlySerTyrGlnGlyTrpLysAsnSerValArgHisAsnLeuSerLeuAsn 61
QY 520 GACTGCTTCGTCAAGTCTGCGGACCCCTCGGCGCCCTGGGCAAGCAACTACTGG 579
Db 62 GlucysPheIleLysLeuProLysGlyLeuGlyArgPro---GlyLysGlyHisTyrTrp 80
QY 580 ATGCTCAACCCCAACAGCGAGTACACTTCGCGCGGGGTCTTCGCGCCGCCCGCAAG 639
Db 81 ThrIleAspProAlaSerGluPheMetPheGluAsnGlySerPheArgArgArgArg 100
RESULT 11
US-09-083-351-15
; Sequence 15, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-083-351-15
```

```
; MOLECULE TYPE: protein
US-09-083-352-15

Alignment Scores:
Pred. No.: 1.05e-12 Length: 106
Score: 332.00 Matches: 62
Percent Similarity: 79.00% Conservatives: 17
Best Local Similarity: 62.00% Mismatches: 19
Query Match: 15.17% Indels: 2
DB: 3 Gaps: 2

US-10-087-080-31 (1-1209) x US-09-083-352-15 (1-106)
QY 343 CGGCGGCC---AAGCCCCCTACTCGTACATCGCGCTCATGCCATGGCCATCCGGAC 399
Db 2 ArgArgProGluLysProProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSer 21
QY 400 TCGCGCGCGCGCGCTGAGCTGGCGGAGATCAACAGTACCTCATGGCGCAAGTTCCCC 459
Db 22 SerProSerLysArgLeuThrLeuSerGluIleTyrGlnPheLeuGlnAlaArgPhePro 41
QY 460 TTTTTCGCGCGAGCTACACGGGCTGGCGCAACTCCGTGGCGCCACAACTTTTCGCTCAAC 519
Db 42 PhePheArgGlyAlaTyrGlnGlyTyrLysAsnSerValArgHisAsnLeuSerLeuAsn 61
QY 520 GACTGCTTCGTAAGTGTGCGGACCCCTCGCGGCGCTGGGCGAAGCAACTACTGG 579
Db 62 GluCysPheIleLysLeuProLysGlyLeuGlyArgPro---GlyLysGlyHisTyrTrp 80
QY 580 ATGCTCAACCCCAACAGCGAGTACACCTTCGCGAGCGGGTCTTCGCGCGCGCGCGCAAG 639
Db 81 ThrIleAspProAlaSerGluPheMetGluAsnGlySerPheArgArgArgArg 100

RESULT 12
US-09-083-352-15
; Sequence 15, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-09-083-352-15

Alignment Scores:
Pred. No.: 1.05e-12 Length: 106
Score: 332.00 Matches: 62
Percent Similarity: 79.00% Conservatives: 17
Best Local Similarity: 62.00% Mismatches: 19
Query Match: 15.17% Indels: 2
DB: 3 Gaps: 2

US-10-087-080-31 (1-1209) x US-09-083-352-15 (1-106)
QY 343 CGGCGGCC---AAGCCCCCTACTCGTACATCGCGCTCATGCCATGGCCATCCGGAC 399
Db 2 ArgArgProGluLysProProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSer 21
QY 400 TCGCGCGCGCGCGCTGAGCTGGCGGAGATCAACAGTACCTCATGGCGCAAGTTCCCC 459
Db 22 SerProSerLysArgLeuThrLeuSerGluIleTyrGlnPheLeuGlnAlaArgPhePro 41
QY 460 TTTTTCGCGCGAGCTACACGGGCTGGCGCAACTCCGTGGCGCCACAACTTTTCGCTCAAC 519
Db 42 PhePheArgGlyAlaTyrGlnGlyTyrLysAsnSerValArgHisAsnLeuSerLeuAsn 61
QY 520 GACTGCTTCGTAAGTGTGCGGACCCCTCGCGGCGCTGGGCGAAGCAACTACTGG 579
Db 62 GluCysPheIleLysLeuProLysGlyLeuGlyArgPro---GlyLysGlyHisTyrTrp 80
QY 580 ATGCTCAACCCCAACAGCGAGTACACCTTCGCGAGCGGGTCTTCGCGCGCGCGCGCAAG 639
Db 81 ThrIleAspProAlaSerGluPheMetGluAsnGlySerPheArgArgArgArg 100

RESULT 13
US-09-087-134-14
; Sequence 14, Application US/09087134
; Patent No. 6365711
; GENERAL INFORMATION:
; APPLICANT: Malcolm Whitman and Xin Chen
; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING
; TITLE OF INVENTION: TGF-BETA SUPERFAMILY SIGNALING
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/087,134
; FILING DATE: 27-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/047,991
; FILING DATE: 28-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bieker-Brady, Kristina
; REGISTRATION NUMBER: 39,109
; REFERENCE/DOCKET NUMBER: 00246/501002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 544 amino acids
; TYPE: amino acid
```



Query Match: 14.64% Indels: 1  
DB: 3 Gaps: 1

US-10-087-080-31 (1-1209) x US-09-083-351-12 (1-106)

QY 340 ACGCGCGGCCCAAGCCCTACTGTCATCGCGCTCATGCCATGCCATCGCGAC 399  
DB 2 ThrArgLeuVallysProPtyrSerTyriAlaLeuileThrMetAlaileuGln 21  
QY 400 TCGCGCGCGCGCTTACGCTGGCGGAGATCAACGAGTACTCATGGCAAGTTCGCC 459  
DB 22 SerProLysLysArgLeuThrLeuSerGluileCysGluPheIleSerGlyArgPhePro 41  
QY 460 TTTTTCGCGCGACGTACACGGCTGGCGCAACTCGCTGGCGCACAACTTTCGCTCAAC 519  
DB 42 TyrTyArgGluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsn 61  
QY 520 GACTGCTTCTCAAGTGTGTCGCGACCCCTCGCGGCGGCAAGGACAACTACTGG 579  
DB 62 AspCysPheVallysIleProArgGluProGlyAsnPro---GlyLysGlyAsnTyTrp 80  
QY 580 ATGCTCAACCCCAACAGCGAGTACACTTCGCGCGCGGGTCTTCGCGCGCGCGCAAG 639  
DB 81 ThrLeuAspProGluSerAlaAspMetPheAspAsnGlySerPheLeuArgArgLys 100  
QY 640 CGCTCAGCCACCGC 654  
DB 101 ArgPheLysArgGln 105

RESULT 15

US-09-083-352-12  
Sequence 12, Application US/09083352  
Patent No. 6207450  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS  
TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/083,352  
APPLICATION NUMBER: US/09/083,352  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-352-12

Alignment Scores:

Pred. No.: 5.09e-12 Length: 106  
Score: 320.50 Matches: 61  
Percent Similarity: 71.43% Conservative: 14  
Best Local Similarity: 58.10% Mismatches: 29  
Query Match: 14.64% Indels: 1  
DB: 3 Gaps: 1

US-10-087-080-31 (1-1209) x US-09-083-352-12 (1-106)

QY 340 ACGCGCGGCCCAAGCCCTACTGTCATCGCGCTCATGCCATGCCATCGCGAC 399  
DB 2 ThrArgLeuVallysProPtyrSerTyriAlaLeuileThrMetAlaileuGln 21  
QY 400 TCGCGCGCGCGCTTACGCTGGCGGAGATCAACGAGTACTCATGGCAAGTTCGCC 459  
DB 22 SerProLysLysArgLeuThrLeuSerGluileCysGluPheIleSerGlyArgPhePro 41  
QY 460 TTTTTCGCGCGACGTACACGGCTGGCGCAACTCGCTGGCGCACAACTTTCGCTCAAC 519  
DB 42 TyrTyArgGluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsn 61  
QY 520 GACTGCTTCTCAAGTGTGTCGCGACCCCTCGCGGCGGCAAGGACAACTACTGG 579  
DB 62 AspCysPheVallysIleProArgGluProGlyAsnPro---GlyLysGlyAsnTyTrp 80  
QY 580 ATGCTCAACCCCAACAGCGAGTACACTTCGCGCGCGGGTCTTCGCGCGCGCGCAAG 639  
DB 81 ThrLeuAspProGluSerAlaAspMetPheAspAsnGlySerPheLeuArgArgLys 100  
QY 640 CGCTCAGCCACCGC 654  
DB 101 ArgPheLysArgGln 105

Search completed: March 29, 2004, 12:05:37  
Job time : 29.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 29, 2004, 12:03:47 ; Search time 59 Seconds  
(without alignments)  
10723.699 Million cell updates/sec

Title: US-10-087-080-31  
Perfect score: 2189  
Sequence: 1 atgaagtggagggtgttgt.....tggagagcctcctagcttga 1209

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 2130338

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlsp  
-MODEL=frame\_n2p\_model -DEV=xlsp  
-Q=/cn2\_1/USPTO\_spool\_p/US10087080/runat\_29032004\_114529\_6667/app\_query.fasta\_1.1351  
-DB=Published Applications AA -QMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10087080 @CGN 1.1.13 @runat\_29032004\_114529\_6667  
-NCPU=6 -ICPU=3 -NO WRAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-FLOGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:  
1: /cn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp.\*  
2: /cn2\_6/ptodata/1/pubpaa/PTCT\_NEW\_PUB.pcp.\*  
3: /cn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pcp.\*  
4: /cn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pcp.\*  
5: /cn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pcp.\*  
6: /cn2\_6/ptodata/1/pubpaa/PTCTUS\_PUBCOMB.pcp.\*  
7: /cn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pcp.\*  
8: /cn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp.\*  
9: /cn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pcp.\*  
10: /cn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pcp.\*  
11: /cn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pcp.\*  
12: /cn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pcp.\*  
13: /cn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pcp.\*  
14: /cn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pcp.\*  
15: /cn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pcp.\*  
16: /cn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pcp.\*  
17: /cn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pcp.\*  
18: /cn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
------------	-------------	--------------	----	-------------

Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:
	3.14e-92	2123.00	100.00%	100.00%	96.98%	402	402	0	0	0

ALIGNMENTS

RESULT 1  
US-10-229-345-18  
; Sequence 18, Application US/10229345  
; Publication No. US20040038220A1  
; GENERAL INFORMATION:  
; APPLICANT: MARKOWITZ, Sanford D.  
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS  
; FILE REFERENCE: CWRU-P01-003  
; CURRENT APPLICATION NUMBER: US/10/229,345  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 18  
; LENGTH: 402  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-229-345-18

1	2123	97.0	402	12	US-10-229-345-18	Sequence 18, Appl
2	2123	97.0	402	12	US-10-274-177-18	Sequence 18, Appl
3	2123	97.0	402	15	US-10-087-080-32	Sequence 32, Appl
4	517	23.6	465	14	US-10-205-823-136	Sequence 136, Appl
5	429.5	19.6	553	10	US-09-292-862-2	Sequence 2, Appl1
6	415.5	19.0	376	14	US-10-007-280A-221	Sequence 221, Appl
7	388.5	17.7	19662	15	US-10-084-846A-6	Sequence 6, Appl1
8	387	17.7	473	9	US-09-844-353A-99	Sequence 99, Appl
9	387	17.7	473	14	US-10-177-293-212	Sequence 212, Appl
10	384.5	17.6	501	9	US-09-963-285-2	Sequence 2, Appl1
11	384.5	17.6	501	9	US-09-963-285-10	Sequence 10, Appl
12	384	17.5	494	9	US-09-963-285-7	Sequence 7, Appl1
13	372.5	17.0	19723	15	US-10-084-846A-5	Sequence 5, Appl1
14	369	16.4	19662	15	US-10-084-846A-6	Sequence 6, Appl1
15	366.5	16.7	237	14	US-10-029-386-32327	Sequence 32327, A
16	364	16.6	19652	15	US-10-084-846A-7	Sequence 7, Appl1
17	354	16.2	19608	15	US-10-084-846A-8	Sequence 8, Appl1
18	351	15.6	19652	15	US-10-084-846A-9	Sequence 9, Appl1
19	339.5	15.5	19695	15	US-10-084-846A-3	Sequence 3, Appl1
20	338.5	15.5	19725	15	US-10-084-846A-4	Sequence 4, Appl1
21	336	15.3	347	9	US-09-844-353A-100	Sequence 100, Appl
22	336	15.0	19608	15	US-10-084-846A-8	Sequence 8, Appl1
23	329	15.0	544	13	US-10-044-442-14	Sequence 14, Appl
24	323.5	14.4	19723	15	US-10-084-846A-5	Sequence 5, Appl1
25	315.5	14.0	19695	15	US-10-084-846A-3	Sequence 3, Appl1
26	306.5	14.0	417	9	US-09-963-285-4	Sequence 4, Appl1
27	305.5	14.0	534	13	US-10-044-442-11	Sequence 11, Appl
28	305	13.9	503	12	US-10-425-114-64093	Sequence 64093, A
29	304.5	13.6	19725	15	US-10-084-846A-4	Sequence 4, Appl1
30	302.5	13.8	777	14	US-10-156-761-10907	Sequence 10907, A
31	301.5	13.4	595	9	US-09-738-973-187	Sequence 187, App
32	301.5	13.4	595	9	US-09-854-133-187	Sequence 187, App
33	301.5	13.4	595	14	US-10-144-649A-187	Sequence 57324, A
34	294	13.4	316	12	US-10-425-114-57324	Sequence 57324, A
35	291	13.0	641	14	US-10-138-098-52	Sequence 52, Appl
36	291	13.0	641	14	US-10-294-804-4	Sequence 36798, A
37	288	13.2	312	12	US-10-425-114-36798	Sequence 5, Appl1
38	283	12.9	960	15	US-10-342-331-5	Sequence 56601, A
39	282.5	12.9	336	12	US-10-425-114-56601	Sequence 57887, A
40	282	12.9	287	12	US-10-425-114-57887	Sequence 14622, A
41	281.5	12.9	375	14	US-10-156-761-14622	Sequence 8, Appl1
42	279.5	12.8	1259	14	US-10-260-715-8	Sequence 47824, A
43	278.5	12.7	389	12	US-10-425-114-47824	Sequence 72062, A
44	278.5	12.7	389	12	US-10-425-114-72062	Sequence 4, Appl1
45	277	12.7	720	15	US-10-342-331-4	



421 CTGGCGGAGATCAACGAGTACTCATGGGCAAGTTCCTCTTTTCCGGCGGAGCTACAG 480  
141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160  
481 GGCTGGCGCAACTCGTGGCGCCCAACCTTCCTCAACGACTCTTCGTCACAGTCTG 540  
161 GlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180  
541 CGGACCCCTCGCGCCCTGGGCAAGGACACTACTGGATGCTCAACCCCAACAGCGAG 600  
181 ArgAspProSerArgProTrpGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200  
601 TACACTTCGCGCGAGCGGTCTTCGCGCGCGCGCCCAAGCGCTCAGCCACCGCGCGCG 660  
201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAlaPro 220  
661 GTCCCGCGCGCGCTCGGCGCGCGAGAGCGCGCGCTCCCGCGCGCGCGCGCGCGCC 720  
221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240  
721 GCGCGCGCGCGCGCTCGCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCC 780  
241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaAlaArgGlnGluAla 260  
781 AGCCCGCGCGCGCAATTCCTCAGCTCTTCGCTGCGATCGACATCTCTCGCAAGCCCTTC 840  
261 SerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280  
841 CGCAGCGCTCGCTCAGGACACAGCGCGCGCGCGCGCTTCAGTGGGGCGCGCGCGCC 900  
281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro 300  
901 TGCCCGCGCGCTCGCGCTTCCTCCCGCTCTCCCGCGCGCGCGCTCGAGGCGCTCGTG 960  
301 CysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeu 320  
961 CGCTCTCGCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGAGTG 1020  
321 ProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluVal 340  
1021 CCACGACCGCGCGCGCTCTCTGCTTGCACCTTCTCCCGCGCGCGCGCGCGCGCGCA 1080  
341 ProProThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaProAlaLysPro 360  
1081 CTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
361 LeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAla 380  
1141 CTGCGCGCGCGCTTGTAGTCGCGCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200  
381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThrLeu 400  
1201 CTAGCT 1206  
401 LeuAla 402

RESULT 3  
US-10-087-080-32  
; Sequence 32, Application US/10087080  
; Publication NO. US20030235820A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Markowitz, Sanford David  
; APPLICANT: Eos Biotechnology, Inc.  
; APPLICANT: Case Western Reserve University  
; TITLE OF INVENTION: No. US20030235820A1el Methods of Diagnosis of Metastatic Colorectal  
; TITLE OF INVENTION: Cancer, Compositions and Methods of Screening for  
; TITLE OF INVENTION: Modulators of Metastatic Colorectal Cancer  
; FILE REFERENCE: 018501-000840US  
; CURRENT APPLICATION NUMBER: US/10/087,080  
; CURRENT FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: US 60/272,206  
; PRIOR FILING DATE: 2001-02-27

; PRIOR APPLICATION NUMBER: US 60/281,149  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: US 60/284,555  
; PRIOR FILING DATE: 2001-04-17  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 402  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: winged helix/forkhead transcription factor (HFH1)  
US-10-087-080-32

Alignment Scores: 3,14e-92 Length: 402  
Pred. No.: 2123.00 Matches: 402  
Score: 2123.00 Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 96.98% Gaps: 0  
DB: 15

US-10-087-080-31 (1-1209) x US-10-087-080-32 (1-402)

QY 1 ATGAAGTTGGAGGTGTCTCTCCCTCGCGCGCGCCACCGGGGCAAGCAGCGAGTACCTG 60  
Db 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
QY 61 GAGGGCGCGCGCGCGCGAGCAGCGCGCTCCCGCTCTCGCGCGCGCGAGACGACTCCCTG 120  
Db 21 GluGlyAlaGlyLysSerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40  
QY 121 GGTCTAGATGGGAGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
Db 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaAlaArgSerThrGln 60  
QY 181 GGGCAGCGGCAACAGAGTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240  
Db 61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIleProAlaAla 80  
QY 241 GCTGCTGAGCGTGTGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 81 AlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100  
QY 301 GGGCGCGGAGCG 360  
Db 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120  
QY 361 TACTCTATCATCGCTCATCGCATCGCGACTCGCGACTCGCGCGCGCGCGCGCGCTGAGC 420  
Db 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyArgLeuThr 140  
QY 421 CTGGCGGAGATCAACGAGTACTCTCATGGCAAGTTCCTCTTTTCCGGCGGAGCTACAG 480  
Db 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160  
QY 481 GGCTGGCGCAACTCTCGTGGCGCCCAACCTTCCTCAACGACTCTTCGTCACAGTCTG 540  
Db 161 GlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180  
QY 541 CGGACCCCTCGCGCCCTGGGCAAGGACACTACTGGATGCTCAACCCCAACAGCGAG 600  
Db 181 ArgAspProSerArgProTrpGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200  
QY 601 TACACTTCGCGCGAGCGGTCTTCGCGCGCGCGCCCAAGCGCTCAGCCACCGCGCGCG 660  
Db 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAlaPro 220  
QY 661 GTCCCGCGCGCGCTCGGCGCGCGAGAGCGCGCGCTCCCGCGCGCGCGCGCGCGCC 720  
Db 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240  
QY 721 GCGCGCGCGCGCGCTCGCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCC 780

```
Db 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluAla 260
Qy 781 AGCCCGGGGGAAGTTCTCCAGCTCCTTCGCCATCGACAGATCTCGCGAAGCCCTTC 840
Db 261 SerProAlaGlyLysPheSerSerPheAlaAlaAspSerPheLeuArgLysProPhe 280
Qy 841 CGCAGCGCTCGCTCAGGACACGGCCCGCGGACGACCTTCAGTGGGGCGCCGCGCC 900
Db 281 ArgSerArgLeuArgAspThrAlaProGlyThrLeuGlnTrpGlyAlaAlaPro 300
Qy 901 TGCCCGCGGTGCGCGGTTCCTCCGCGCTCTCCCGCGCGCGCTTCGAGGCCCTGCTG 960
Db 301 CysProLeuProAlaPheProAlaLeuLeuProAlaProCysArgAlaLeuLeu 320
Qy 961 CCCTCTGCGCGACGGCGCGGCGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGG 1020
Db 321 ProLeuGlyAlaGlyGlyAlaGlyGlyProAlaArgLeuGlyAlaArgGluAla 340
Qy 1021 CCACCGACCGCGCGCGCTCTGCTTGTGACCTCTCCCGCGCGCGCGCGCGCGCG 1080
Db 341 ProProThrAlaProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysPro 360
Qy 1081 CTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Db 361 LeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyCysProLeuArgLeuProAlaAla 380
Qy 1141 CTCGAGCGCGCTTAGTCCGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCG 1200
Db 381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyProValGluThrLeu 400
Qy 1201 CTAGCT 1206
Db 401 LeuAla 402

RESULT 4
US-10-205-823-136
; Sequence 136 Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR FILING DATE: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-10-205-823-136
Alignment Scores:
Pred. No.: 6,37e-17 Length: 465
Score: 517.00 Matches: 157
Percent Similarity: 44.21% Conservative: 34
Best Local Similarity: 36.34% Mismatches: 132
Query Match: 23.62% Indels: 109
DB: 14 Gaps: 16

US-10-087-080-31 (1-1209) x US-10-205-823-136 (1-465)
Qy 60 GGAGGCGCGGGCGGCGGAGCGAGCGCGCTCCCGCTTCGCGCGGGGAGCGACTCCCT 119
Db 39 GlyGlyGlyGlyGlyProArgLeuAlaValPro----- 49
Qy 120 GGGCTCAGATGGGACTGCGCGGCAAGCGCTCGCGGGGGGGCGGCGCGGACAGATACGCA 179
Db 50 -----AlaGlnArgArgArgArgSerTyAla 60
Qy 180 GGGCGAGCGGCAACAGAG-----TGCGGG 203
Db 61 GlyGluAspGluLeuGluAspLeuGluGluGluAspAspAspAspPileLeuLeuAla 80
Qy 204 AGGCGGCGCGGCGCGGAGGCGATCCCGGCGAGCGCTGCTCGACGCGGTGTGTGGCGGA 263
Db 81 ProProAlaGlyGlySerProAlaProProGlyProAlaProAlaAlaGlyAlaGlyAla 100
Qy 264 GGGCGCGAGCGCGGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 323
Db 101 GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 119
Qy 324 ACGCAGCAAGCCATATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 383
Db 119 aLysAsn---ProLeuVal-----LysProProTySerTyTrpLeuAlaLeuLeuLeu 135
Qy 384 CATGCCATCCGCGACTCGGCGGCGGCGGCTTCACTGCGCGGAGATCAACGAGTACTT 443
Db 135 rMetAlaAlaLeuGlnSerProLysLysArgLeuThrLeuSerGluLeuLeuCysGluPheIl 155
Qy 444 CATGGGCAAGTTCCCTTTTCGCGGCGAGCTACACGGGCTGGCGCAACTCCGCGCGCA 503
Db 155 eSerGlyArgPheProTyTyArgGlyLysPheProAlaTrpGlnAsnSerIleArgHi 175
Qy 504 CAACCTTTTCGCTCAACGACTGCTTCGTCAGGTGCTCGCGACCCCTCGCGGCGCGTGG 563
Db 175 sAnLeuSerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro---Gl 194
Qy 564 CAAGGCAACTACTGATGCTCAACCCCAACGAGGAGTACACCTTCGCGCGGCGGCTCTT 623
Db 194 yLysGlyAsnTyTrpThrLeuAspProGluSerAlaAspMetPheAspAsnGlySerPh 214
Qy 624 CGCGCGCGCGCAAGCGCTCAGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCG 683
Db 214 eLeuArgArgArgLysArgPheLys---ArgGlnProLeuLeuProProAsnAlaAla 233
Qy 684 CGAGGAGGCC-----CCGGGCTCCCGCGCG 710
Db 233 aAlaGluSerLeuLeuLeuArgGlyAlaGlyAlaAlaGlyAlaGlyAlaAspProAlaAl 253
Qy 711 CCGG-----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGC----- 742
Db 253 aAlaAlaAlaLeuPheProProAla---ProProProProProHisAlaTyGlyTyGlyP 273
Qy 743 -----CCGGATGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGGA 773
Db 273 rOTyRGlyCysGlyTyGlyLeuGlnLeuProProTyTyAlaProProSerAlaLeuPheA 293
Qy 774 GCGCGCGAGCGCGCGGCGCAAGTTCTCCAGTCTCTCGCCATCGACAGCATCTCGCGCA 833
Db 293 lAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaPheHisProHis-----S 308
Qy 834 GCCCTTCGCGAGCGCTCGCTCAGGACACAGCGCGCGCGCGCGCGCGCTTCAGTGGGCGC 893
```

[illegible]

## RESULT 5

```

US-09-292-862-2
; Sequence 2, Application US/09292862
; Publication NO. US20030013087A1
; GENERAL INFORMATION:
; APPLICANT: Walter, Michael A.
; APPLICANT: Jordan, Tim
; APPLICANT: Raymond, Vincent
; TITLE OF INVENTION: NOVEL MUTATIONS IN THE FRAC3 GENE FOR
; TITLE OF INVENTION: DIAGNOSIS AND PROGNOSIS OF GLAUCOMA AND ANTERIOR SEGMENT
; TITLE OF INVENTION: DYSGENESIS
; FILE REFERENCE: 07540/020003
; CURRENT APPLICATION NUMBER: US/09/292,862
; CURRENT FILING DATE: 1999-04-16
; EARLIER APPLICATION NUMBER: 60/084,784
; EARLIER FILING DATE: 1998-05-08
; EARLIER APPLICATION NUMBER: 60/082,206
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-292-862-2

```

Alignment Scores:	7.87e-13	Length:	553
Pred. No.:	493.50	Matches:	146
Score:	493.50	Conservative:	38
Percent Similarity:	41.82%	Mismatches:	124
Best Local Similarity:	33.18%	Indels:	132
Query Match:	19.62%	Gaps:	18
DB:	10		

US-10-087-080-31 (1-1209) x US-09-292-862-2 (1-553)

Qy	214	GGCGGGAGGAGCGCATCCCGACGACGACTGCTGCACGGTGTGCGGAGGCGCGGAG	273
Db	20	GlyGlyGlulnSerTyrTyrArgAlaAlaAlaAla-----AlaGlyGlyGlyTyr	37
Qy	274	GCGGGGCGGGCGGCGGCGGCGG-----	297
Db	38	ThrAlaMetProAlaProMetSerValTyrSerHisProAlaHisAlaGlulnTyrPro	57
Qy	298	GGCGGGCGGGAGGCGGCGGCGGTCACGACGAAGCCATATACGCGGGGGCC-----	351

Dd		58 GlyGlyMetAlaArgAlaTyrgLy-----ProTyrThrProGlnProGlnPro 73
Qy	352	-----AAGCCCCCTACTCGTACATCGCGCTCATCCCAATGGCCATTCGGCAG 399
Dd		74 LysAspMetValIysProTyrSerTyrIleAlaLeuIleThrMetAlaIleGlnAasn 93
Qy	400	TGCGCGGGCGGGCGCTTGACGTGCGGAGATCAACAGAGTACTCATGGGCAGATTCCCC 459
Dd		94 AlaProAspLysIleThrLeuAanGlyIleTyrGlnPheIleMetAspArgPhePro 113
Qy	460	TTTTTCCGGCGCATACACGGGCTCGGGAATCCGTGCGGCCACAACTTTTCGTCAAC 519
Dd		114 PheTyrArgAspAsnIysGlnGlyTyrPginAasSerIleArgHisAenLeuSerLeuAasn 133
Qy	520	GACTGCTTCGTCAAGTGCTGCGCACCTCCGCGCCCTCGCGCCGAGGACAATACTCTGG 579
Dd		134 GluCysPheValIysValProArgAspAspLysIlePro----GlyIysGlySerTyrTip 152
Qy	580	ATGCTCAACCCCACAGCAGTAGTACACTTCGCGCAGCGGGTCTTCGCGCCGCGCGCAAG 639
Dd		153 ThrLeuAspProAspSertYrAenMetPheGluAanGlySerPheLeuAArgAtgArg 172
Qy	640	CQCCTCAGC-----C-----CACC GCGCG 657
Dd		173 ArgPheLysLysAspAlaLeuLysAspLysGluGluLysAspArgLeuHisLeuLys 192
Qy	658	CCGCTCCCGCGCCCGGGTCCGCGCCGAGAGCGCCCGCGCTCCCGCGC----- 708
Dd		193 GluProProProGlyAlaSerProArgProAlaProProGluGlnAlaAspGlyAasn 212
Qy	709	GCCCGCGCGCCGCGCGCGCGC-----CGGCCTCGCGCGC 747
Dd		213 AlaPro-GlyProGlnProProProValArgIleGlnAspIleLysThrGluAanGlyTh 232
Qy	748	ATCGCTCGCCCGCCCGCAGGAGGAGC---GGCCAGCCCGCGGCAAGTTCCTCAGC 804
Dd		232 rCysProSerProProGlnProLeuSerProAlaAlaAlaLeuGlySerGlySerAlaAl 252
Qy	805	TCCTTCGCCATCGACAGCATCTCGCGAACCCCTTCGCGACCGCTCCCTCAGGACAG 864
Dd		252 aaIaValProLysIleGluSerProAspSerSerSerSerLeuSerGlySerSe 272
Qy	865	GCCCCCGGGA-----CGAGCTTCAGT-----GGGGCGCC 894
Dd		272 rP-ProGlySerLeuProSerAlaArgProLeuSerLeuAspGlyAlaAspSerAlaPr 292
Qy	895	GCGCCCTCGCCGCGCTCGCCGCGTTCGCCGCGCTCTCC----- 935
Dd		292 oProPro-ProAla-----ProSerAlaProProChIeHisSerGlnGlyP 308
Qy	936	-----CGCGCGCCCTGCAGGGCCCTGCTCGCCGCTC 966
Dd		308 heSerValAspAenIleMetThrSerLeuArgGlySerProGlnSerAlaAlaIadLu 328
Qy	967	TGCGCGTACGCGCGCGAGCGCGCGCTCGCGCGCGAGCGCGAGTGCACCG 1026
Dd		328 eu-----SerSerGlyLeuLeuAlaSerAlaAlaSerS 340
Qy	1027	ACCGCGGC-----GCCCTTCCT----- 1043
Dd		340 erArgAlaGlyIleAlaProProLeuAlaLeuGlyAlaTyrSerProGlyInSerSerL 360
Qy	1044	-----CCTTCACCTCTCCCGGGGGCGCCCGCC 1074
Dd		360 euTyrSerSerProCysSerGlnThrSerSerAlaGlySerSerGlyGlyGlyGlyG 380
Qy	1075	AAGCCACTCGAGCCCGCGCGCGCGCGGACCTGTAC---TGCCTCTCGG--- 1128
Dd		380 lyAlaGlyAlaAlaGlyGly-AlaGlyAlaGlyThrTyrHisCysAenLeuGlnAla 399
Qy	1129	-----CTGCGCGCAGCCCTG 1143
Dd		400 MetSerLeuTyrAlaAlaGlyGluArgGlyGlyHisLeuGlnGlyAlaProGlyGlyAla 419

[illegible]



Db 107 -----AlaLeuSerProSer---GlyMetGlyAlaMetGlyAlaGlnGlnAla 122  
Qy 190 GAACAGAGTGGCGGAGCGCGCGCGAGAGCGATCCCGCAGCAGCTGCTGCA 249  
Db 123 SerMetMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCysMetSerProMet 142  
Qy 250 GCGGTGGTGGCGGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309  
Db 143 AlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGly-----GlyGlyGlyAsp 159  
Qy 310 AGCGGCGAGGTGTCAGCGAGAGCATATACCGCGCGCGCGCGCGCGCGCGCG 369  
Db 160 AlaLysThrPhenylsArgSerTyrProHisAla-----LysProProTyrSerTyr 176  
Qy 370 ATCGGCTCATCCGTCGTCCTCCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 429  
Db 177 IleSerLeuIleThrMetAlaIleGlnArgAlaProSerLysMetLeuThrLeuSerGlu 196  
Qy 430 ATCAACGAGTACCTATGGCAAGTTCCTCTTTCGCGGAGCTACACGGCTGGCG 489  
Db 197 IleTyrGlnTrpIleMetAspLeuPheProTyrTyrArgGlnAsnGlnArgTrpGln 216  
Qy 490 AACTCGGTGCGCACACCTTCCTCAAGCTCTGCTCAAGGTGCTGCGGACCC 549  
Db 217 AsnSerIleArgHisSerLeuSerPheAsnAspCysPheValysValalaArgSerPro 236  
Qy 550 TCGCGCCCTGGCGGAGCACTACTGATGCTCAACCCCAACAGCAGTACACCTTC 609  
Db 237 AspLysPro---GlyLysGlySerTyrTrpThrLeuHisProAspSerGlyAsnMetPhe 255  
Qy 610 GCGAGCGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669  
Db 256 GluAsnGlyCysTyrLeuArgArgGlnLysArg----- 266  
Qy 670 CCGCGGTGCGCGCGAGAGAGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCG 729  
Db 267 -----PhelyCysGlnLysGlnProGly-----AlaGlyGly 277  
Qy 730 GCGCGGCTGCGCGCGCGATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 783  
Db 278 GlyGlyGlySerGlySerGlyGlyAlaLysGlyGlyProGluSerArgLysAsp 297  
Qy 784 CCGCGGGCAAGTCTCCAGCTCTTCGCGATCGACGATCTCGCGCAAGCTTCGCG 843  
Db 298 ProSerGly-----AlaSerAsnProSerAlaAspSerProLeuHisArgGlyValHis 315  
Qy 844 AGCGGTGCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903  
Db 316 -----GlyLysThrGlyGlnLeuGluGlyAlaProAla 326  
Qy 904 CCGCGGTGCGCGCGGTTCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 963  
Db 327 ProGlyProAlaAlaSerProGlnThrLeuAsp----- 337  
Qy 964 CTCTGCGGTGACG 1023  
Db 338 -----HisSerGlyAlaThrAlaThrGlyGlyAlaSerGluLeuLysThrPro 353  
Qy 1024 -----CCGAGCGCGCGCGCGCTC-----CTGCTGCTGCTCCTCGCGCG 1062  
Db 354 AlaSerSerThrAlaProProIleSerSerGlyProGlyAlaLeuAlaSerValProAla 373  
Qy 1063 GCG 1122  
Db 374 SerHisProAlaHisGlyLeu---AlaProHisGluSerGlnLeuHisLeu----- 389  
Qy 1123 CTGCGGTGCG 1182  
Db 390 -----LysGlyAspProHisTyrSer 396  
1183 TAC 1185  
Qy  
397 Phe 397

RESULT 9  
US-10-177-293-212  
; Sequence 212, Application US/10177293  
; Publication No. US20030124128A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Gannavarpu, Manjula  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Mertens, Maureen  
; APPLICANT: Myer, Vic  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Monahan, John  
; APPLICANT: Meyers, Rachel E.  
; APPLICANT: Bast Jr., Robert C.  
; APPLICANT: Hortobagyi, Gabriel N.  
; APPLICANT: Puztai, Lajos  
; APPLICANT: Mexic, Funda  
; APPLICANT: Sahin, Aysegul  
; APPLICANT: Mills, Gordon B.  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-038  
; CURRENT APPLICATION NUMBER: US/10177,293  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: US 60/299,887  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 60/301,572  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: US 60/306,501  
; PRIOR FILING DATE: 2001-07-18  
; PRIOR APPLICATION NUMBER: US 60/325,002  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US 60/362,585  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
; PRIOR FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 506  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 212  
; LENGTH: 473  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-177-293-212  
  
Alignment Scores:  
Pred. No.: 7,93e-11 Length: 473  
Score: 387.00 Matches: 124  
Percent Similarity: 44.36% Conservative: 45  
Best Local Similarity: 32.55% Mismatches: 132  
Query Match: 17.68% Indels: 80  
DB: 14 Gaps: 15  
  
US-10-087-080-31 (1-1209) x US-10-177-293-212 (1-473)  
Qy 70 GCGCGCAGCAGCGCGCGCTCCCGCTGTCGCGCGCGGAGCAGCAGCTCCCTGGGCTCAGAT 129  
Db 88 GlySerAlaGlyAlaMetAsnSerMetThrAlaAlaGlyValThrAlaMetGlyThr--- 106  
Qy 130 GGGGACTGCGCGGCAAGCCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 189  
Db 107 -----AlaLeuSerProSer---GlyMetGlyAlaMetGlyAlaGlnGlnAla 122  
Qy 190 GAACAGAGTGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 249  
Db 123 SerMetMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCysMetSerProMet 142  
Qy 250 GCGGTGGTGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309

```
Db 143 AlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGly-----GlyGlyGlyAsp 159
Qy 310 ACCGCGAGGTGCGACGAGCAAGCATATACGCGCGGCCCAAGCCCTCTACTCGTAC 369
Db 160 AlaLysThrPheLysArgSerTyrProHisAla-----LysProTyrSerTyr 176
Qy 370 ATCGCGCTCATCGCATCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 429
Db 177 IleSerLeuIleThrMetAlaIleGlnArgAlaProSerLysMetLeuThrLeuSerGlu 196
Qy 430 ATCAACGAGTACTCATGCGGCAAGTTCCTCTTTTCCGCGGAGCAGTACACGGGTGGCG 489
Db 197 IleTyrGlnTrpIleMetAspLeuPheProTyrTyrArgGlnAsnGlnInArgTrpGln 216
Qy 490 AACTCGCTCGCGCCCAACCTTCCTCGTCAAGCTCTTCGTCACAGGTCTCGCGGACCC 549
Db 217 AsnSerIleArgHisSerLeuSerPheAsnAspCysPheValLysValAlaArgSerPro 236
Qy 550 TCGCGGCTCGGGCAAGCAACTACTGATGTCACCAACCCCAACAGCAGTACACCTTC 609
Db 237 AspLysPro---GlyLysGlySerTyrTrpThrLeuHisProAspSerGlyAsnMetPhe 255
Qy 610 GCGGAGGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
Db 256 GluAsnGlyCysTyrLeuArgArgGlnLysArg----- 266
Qy 670 CCGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
Db 267 -----PheLysCysGluLysGlnProGly-----AlaGlyGly 277
Qy 730 GCCCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783
Db 278 GlyGlyLysGlySerGlyGlySerGlyAlaLysGlyGlyProGluSerArgLysAsp 297
Qy 784 CCGCGGCGCAAGTTCCTCGAGTCTCTCGCCATCGACAGCATCTCTCGCAAGCCCTTCG 843
Db 298 ProSerGly-----AlaSerAsnProSerAlaAspSerProLeuHisArgLysValHis 315
Qy 844 AGCGTGGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
Db 316 -----GlyLysThrGlyGlnLeuGluGlyAlaProAla 326
Qy 904 CCGCGGCTCGCGCGTTCCTCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 963
Db 327 ProGlyProAlaAlaSerProGlnThrLeuAsp----- 337
Qy 964 CTCTCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1023
Db 338 -----HisSerGlyAlaThrAlaThrGlyGlyAlaSerGluLeuLysThrPro 353
Qy 1024 -----CGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
Db 354 AlaSerSerThrAlaProProIleSerSerGlyProGlyAlaLeuAlaSerValProAla 373
Qy 1063 GCGGCGCGCGCGCAAGCACTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
Db 374 SerHisProAlaHisGlyLeu---AlaProHisGluSerGlnLeuHisLeu----- 389
Qy 1123 CTGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182
Db 390 -----LysGlyAspProHisTyrSer 396
Qy 1183 TAC 1185
Db 397 Phe 397
```

## RESULT 10

```
US-09-963-285-2
; Sequence 2, Application US/09963285
; Patent No. US2002090707A1
; GENERAL INFORMATION:
; APPLICANT: Enerback, Sven
; APPLICANT: Krook, Katarina
```

```
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-963-285-2

Alignment Scores:
Pred. No.: 1,03e-10 Length: 501
Score: 384.50 Matches: 127
Percent Similarity: 45.86% Conservative: 39
Best Local Similarity: 35.08% Mismatches: 118
Query Match: 17.57% Indels: 78
DB: 9 Gaps: 15

US-10-087-080-31 (1-1209) x US-09-963-285-2 (1-501)
Qy 259 GCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309
Db 27 AlaAlaGlySerTyrGlyGlyMetAlaSerProMetGlyValTyrSerGlyHisProGlu 46
Qy 310 -----AGCGCGAGGTGCGACGAGCAAGCCATATACGCGCGCGCGCGCG 351
Db 47 GlnTyrSerAlaGlyMetGlyArgSerTyrAlaProTyrHisHisGlnProAlaAla 66
Qy 352 -----AAGCCCCCTACTCGTACATCGCGCTCATCGCGCGCGCGCGCGCG 396
Db 67 ProLysAspLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleGln 86
Qy 397 GACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 456
Db 87 AsnAlaProGluLysIleThrLeuAsnGlyIleTyrGlnPheIleMetAspArgPhe 106
Qy 457 CCCTTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
Db 107 ProPheTyrArgGluAsnLysGlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeu 126
Qy 517 AAGCACTGCTTCGTCAGGTGCTGCGCGACCCCTCGCGCGCGCGCGCGCGCGCG 576
Db 127 AsnGluCysPheValLysValProAspAspLysLysPro---GlyLysGlySerTyr 145
Qy 577 TGGATGTCACCCCAACAGCAGTACACCTTCGCGAGCGGGTCTTCGCGCGCGCG 636
Db 146 TrpThrLeuAspProAspSerTyrAsnMetPheGluAsnGlySerPheLeuArgArg 165
Qy 637 AAGCGCTCAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 696
Db 166 ArgArgPheLysLysLys-----AspValSerLysGluLysGluArgAla 181
Qy 697 GGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Db 182 HisLeuLysGluProProProAlaAlaSerLysGlyAlaProAlaThrProHisLeuAla 201
Qy 754 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789
Db 202 AspAlaProLysGluAlaGluLysValIleLysSerGluAlaAlaSerProAla 221
Qy 790 -----GGCAAGTTCCTCAGCTCCTTCGCCATCGACGCA----- 823
Db 222 LeuProValIleThrLysValGluThrLeu-SerProGluSerAlaLeuGlnGlySerPr 241
```

```

QY 824 ----TCTGCGAAGCCCTTCCGACCGCTGCTCAGGACACGGCCCGCGGACGACG 879
Db 241 oArgSerAlaAlaSer-ThrProAlaGlySerProAspGlySerLeuProGluHisA 261
QY 880 CTTTCAGTGGGCGCGCGCCCTCCGCGCGTCC-----CGCGTTCCCG 924
Db 261 laAla---AlaProAsnGlyLeuProGlyPheSerValGluAsnLeuMetThrLeuArgT 280
QY 925 GCCTCTCTCCCGCGCGCGCTGCTCAGGCGCTGCTCCGCTCTGCGCGTACGCGCGCGG 984
Db 280 hrSerProGlyGlyGluLeu-----287
QY 985 GAGCGCGCGCGTGGCGCGCGCGAGCGGAGTCCACGACCGCGCGCGCCCTCTCTG 1044
Db 288 --SerProGlyAlaGlyArgAla-GlyLeuValValProLeuAlaLeuProTyrAla 306
QY 1045 CTTGCACTCTCCCGCGCGCGCGCTGCTCAGGCGCGCTGCTCCGCTCTGCGCGTACGCGCGG 1104
Db 307 AlaAlaPro---ProAlaAlaTyrGlyGlnProCysAlaGlnGlyLeuGluAlaGlyAla 325
QY 1105 GCGCACTGTAC---TGCCCGCTCGCG-----1128
Db 326 AlaGlyGlyTyrGlnCysSerMetArgAlaMetSerLeuTyrThrGlyAlaGluArgPro 345
QY 1129 -----CTGCGCGCGCGCGCTGCTCAGGCGCGCTTGTAGTCGCGCTCTCT 1173
Db 346 AlaHisMetCysValProProAlaLeuAspGluAlaLeuSerAspHisProSerGlyPro 365

RESULT 11
US-09-963-285-10
; Sequence 10, Application US/09963285
; Patent No. US20020090707A1
; GENERAL INFORMATION:
; APPLICANT: Enerbck, Sven
; APPLICANT: Krook, Katarina
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 501
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-963-285-10

Alignment Scores:
Pred. No.: 1.03e-10 Length: 501
Score: 384.50 Matches: 127
Percent Similarity: 45.86% Conservative: 39
Best Local Similarity: 35.08% Mismatches: 118
Query Match: 17.57% Indels: 78
DB: 9 Gaps: 15

```

US-10-087-080-31 (1-1209) x US-09-963-285-10 (1-501)

```

QY 259 GCGGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 309
Db 27 AlaAlaGlySerTyrGlyGlyMetAlaSerProMetGlyValTyrSerGlyHisProGlu 46
QY 310 -----ACGCGCGAGGTGCAGGACGACGACCATATACGCGCGCGCGCC 351
Db 47 GlnTyrSerAlaGlyMetGlyArgSerTyrAlaProTyrHisHisGlnProAlaAla 66

```

# RESULT 12

US-09-963-285-7  
; Sequence 7, Application US/09963285  
; Patent No. US20020090707A1  
; GENERAL INFORMATION:  
; APPLICANT: Enerbck, Sven  
; APPLICANT: Krook, Katarina  
; APPLICANT: Rondahl, Lena

```

QY 352 -----AAGCCCGCTACTCTGATACGCGCTCATCGCATCGGCGCATCGCG 396
Db 67 ProLysAspLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleGln 86
QY 397 GACTCGCGCGCGCGCGCTTACGCTCGCGGAGATCAACGAGTACCTCATCGGCAAGTTC 456
Db 87 AsnAlaProGluLysLysIleThrLeuAsnGlyIleTyrGlnPheIleMetAspArgPhe 106
QY 457 CCCTTTTTCGCGCGCGCTACACGCGTGGCGGCACTCCGTCGCGCACCACTTTCGCTC 516
Db 107 ProPheTyrArgGluAsnLysGlnGlyTyrPheGlnAsnSerIleArgHisAsnLeuSerLeu 126
QY 517 AACCACTGCTTCGTCAGGTGCTCGCGGACCCCTCGCGGCGCTCGGCGCAAGCAACTAC 576
Db 127 AsnGluCysPheValLysValProArgAspAspLysPro---GlyLysGlySerTyr 145
QY 577 TGGATGCTCAACCCCAACAGCGAGTACACCTTTCGCGCGCGGCTTTCGCGCGCGCGCG 636
Db 146 TrpThrLeuAspProAspSerTyrAsnMetPheGluAsnGlySerPheLeuArgArgArg 165
QY 637 AAGCGCTCAGCCACCGCGCGCGTCCCGCGCGCGTTCGCGCGCGCGCGCGCGCGCG 696
Db 166 ArgArgPheLysLysLys-----AspValSerLysGluLysGluGluArgAla 181
QY 697 GCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Db 182 HisLeuLysGluProProProAlaAlaSerLysGlyAlaProAlaThrProHisLeuAla 201
QY 754 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789
Db 202 AspAlaProLysGluAlaGluLysValValIleLysSerGluAlaAlaSerProAla 221
QY 790 -----GGCAAGTTCTCCAGTCTCCATCGCTTCCCATCGACAGCA-----823
Db 222 LeuProValIleThrLysValGluThrLeu-SerProGluSerAlaLeuGlnGlySerPr 241
QY 824 ----TCCTGCGCAAGCCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 879
Db 241 oArgSerAlaAlaSer-ThrProAlaGlySerProAspGlySerLeuProGluHisA 261
QY 880 CTTTCAGTGGGCGCGCGCGCGCTGCTCCGCGCGCGCTGCTCCGCTCTGCGCGTACGCGCGG 924
Db 261 laAla---AlaProAsnGlyLeuProGlyPheSerValGluAsnLeuMetThrLeuArgT 280
QY 925 GCGTCTCTCCCGCGCGCGCGCTGCTCAGGCGCGCTGCTGCGCTCTGCGCGTACGCGCGG 984
Db 280 hrSerProGlyGlyGluLeu-----287
QY 985 GAGCGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTCTG 1044
Db 288 --SerProGlyAlaGlyArgAla-GlyLeuValValProLeuAlaLeuProTyrAla 306
QY 1045 CTTGCACTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1104
Db 307 AlaAlaPro---ProAlaAlaTyrGlyGlnProCysAlaGlnGlyLeuGluAlaGlyAla 325
QY 1105 GCGCACTGTAC---TGCCCGCTCGCG-----1128
Db 326 AlaGlyGlyTyrGlnCysSerMetArgAlaMetSerLeuTyrThrGlyAlaGluArgPro 345
QY 1129 -----CTGCGCGCGCGCGCTGCTCAGGCGCGCTTGTAGTCGCGCTCTCT 1173
Db 346 AlaHisMetCysValProProAlaLeuAspGluAlaLeuSerAspHisProSerGlyPro 365

```

```

; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-963-285-7

Alignment Scores:
Pred. No.: 1,098-10 Length: 494
Score: 384.00 Matches: 129
Percent Similarity: 45.41% Conservative: 39
Best Local Similarity: 34.86% Mismatches: 107
Query Match: 17.54% Indels: 95
DB: 9 Gaps: 19

US-10-087-080-31 (1-1209) x US-09-963-285-7 (1-494)
QY 259 GCGGAGGCGGCGAGCGCGCGCGCGGCGCA---GCGCGGCGCGCGCGCG 306
DB 27 AlaAlaGlySerTyrGlyGlyMetAlaSerProMetGlyValTyrSerGlyHisProGlu 46
QY 307 -----GGAGGCGGAGGCGTGCACGACGAGCATATACGCGCGCGCC- 351
DB 47 GlnTyrGlyAlaGlyMetGlyArgSerTyrAlaProTyrHisGlnProAlaAlaPro 66
QY 352 -----AAGCCCGCTACTCGTACATCGCGTCAATCGCATCGCATCGCGAC 399
DB 67 LysAspLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleGluAsn 86
QY 400 TCGGCGGCGCGCGCTTACGCTGCGGAGATCAACGAGTACCTCATGGCAAGTCCCG 459
DB 87 AlaProGluLysLysIleThrLeuAsnGlyIleTyrGlnPheIleMetAspArgPhePro 106
QY 460 TTTTTCGCGGCGAGCTACACGCGGCTGGCGCAACTCGTGCACCAACTTTTCGCTCAAC 519
DB 107 PheTyrArgGluAsnLysGlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsn 126
QY 520 GACTGCTTCGTCACAGTGTGCGACCCCTCGCGCGCTGGGCGACGAGCAACTACTGG 579
DB 127 GluCysPheValLysValProArgAspLysLysPro---GlyLysGlySerTyrTrp 145
QY 580 ATGCTCAACCCCAACAGCGAGTACACCTTCGCGGAGCGGCTTCGCGCGCGCGCGCAAG 639
DB 146 ThrLeuAspProAspSerTyrAsnMetPheGluAsnGlySerPheLeuArgArgArg 165
QY 640 GCGCTCAGCCACCGCGCGCGTCCCG- 666
DB 166 ArgPheLys---LysLysAspValProLysAspLysGluArgAlaHisLeuLysGlu 184
QY 667 GCGCGCGGCTGCGCGCGGAGGAGCGCGCG---GCGCTCCCGCGCGCGCGCGCGCGCG 720
DB 185 ProProSerThrAlaLysGlyAlaProThrGlyThrProValAlaAspGlyProLys 204
QY 721 -----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 741
DB 205 GluAlaGluLysValValLysSerGluAlaAlaSerProAlaLeuProValIle 224
QY 742 CCCGCGATCGCTCGCGCGCGCGCGGAGGAGCGCGCGAG-----CCCGCGCGCGCAAG 795
DB 225 ThrLysValGluThrLeuSer---ProGluGlyAlaLeuGlnAlaSerProArgSer 242

```

```

QY 796 TTCTCCAGCTCTCTCCCATCGACAGCATCTGCGCAAGCCCTTCCGACGCCCTGCTC 855
DB 243 -----AlaSerSerThrProAlaGlySerProA 252
QY 856 AGGACACAGCCCGCGGAGCGAGCTTTCAGTGGGCGCGCGCGCTGCGCGCGCGCTGCC- 914
DB 252 spGlySerLeuProGluHisAlaAla---AlaProAsnGlyLeuProGlyPheSerV 271
QY 915 -----CGGTTCCCGCGCTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
DB 271 alGluThrIleMetThrLeuArgThrSerProGlyGlyAspLeu---SerProAlaA 290
QY 961 CCCTCTGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
DB 290 laAla-----ArgAlaGlyLeu-Val-----Val 297
QY 1021 CCACCGACGCGCGCGCGCGCTCTCTCTGTCACCTCTCCCGCGCGCGCGCGCGCGCG 1080
DB 298 ProProLeuAlaLeuProTyrAlaAlaAlaPro---ProAlaAlaTyrThrGlnProCys 316
QY 1081 CTCGAGGCG---CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1128
DB 317 AlaGlnGlyLeuGluAlaAlaGlySerAlaGlyTyrGlnCysSerMetArgAlaMetSer 336
QY 1129 -----CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1152
DB 337 LeuTyrThrGlyAlaGluArgProAlaHisValCysValProProAlaLeuAspGluAla 356
QY 1153 TTAGTCGCGCGTCT---GCGCGG 1173
DB 357 LeuSerAspHisProSerGlyPro 364

RESULT 13
US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLERWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 19723
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

Alignment Scores:
Pred. No.: 2,166-10 Length: 19723
Score: 372.50 Matches: 172
Percent Similarity: 35.29% Conservative: 38
Best Local Similarity: 28.91% Mismatches: 162
Query Match: 17.02% Indels: 223
DB: 15 Gaps: 34

US-10-087-080-31 (1-1209) x US-10-084-846A-5 (1-19723)
QY 8 TGGAGGTGTTTCGTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67
DB 6007 TrpArgCysThrProArgArgAlaCysGlySerAlaSerAla---ProTrpA-gTtp 6025

```

QY	68	CGGCGCGCAGCGAGCGCGCGCTCCCGCTGTCGGCGGGGAGACGACTCCCTCGGCTCAG	127
Db	6026	ProGlyAlaSerThr-----CysTrpCysHisArgAlaArgSerProTrpAlaArg	6043
QY	128	-----ATGGGACTGCGCGCCCAAGCCGTCGCGGGCGGGCGGCG	166
Db	6044	ProAlaArgTrpArgProThrAlaGlyAlaSerArgSerProArgArgProAspAlaThr	6063
QY	167	CCAGAGATACGACGGCGACGGCGAACAGAGTGCGGAGCGGGCGCG-----GCGCGG	220
Db	6064	ProLeuProArgArgProThrAlaThrAlaAlaAlaAlaAlaCysTrpSerSer	6083
QY	221	AGAGGGGATCCCGGCAGCAGCTGCTCCACGGGTGTGGCGG-----AGG	265
Db	6084	SerCysArgThrProTrpAlaThrValThrGlyCysTrpArgCysAlaAlaAlaArg	6103
QY	266	GCGCGGAGGCGGGCGCG-----CGGGCCAGGCGCGGGCGGGCGG	304
Db	6104	SerThrArgThrGlyAlaProThrAlaSerTrpProAlaGlyArgProArgSerThr	6123
QY	305	CGGGAGCGGGAGGGTGCAGGACA-----AGCCATATACGGCG	346
Db	6124	SerGlyAlaProArgProArgAlaSerLeuProThrProLeuThrSerSerArgProThr	6143
QY	347	GGGCCAAGCCCC-----CCTACTCGTACATCGCGCTCATCGCCATCGCCATCC	394
Db	6144	AlaProValProAlaSerAlaThrProTrpArgSerAlaArgSerProProSerThrAsp	6163
QY	395	GGCACTCGG-----CGGGCGGGCGCTTGACGTGGCGGAGATCAACGAGTACTCATGG	448
Db	6164	ValThrAlaProValArgSerArgAlaArgSerGlyArgSerProThrSerGlyIleTrp	6183
QY	449	-----GCAAGTTCGCCCTTTTCCGCGCGCAGCTACACGGGCTGGC	487
Db	6184	ArgGlyProProGlySerProProGlySerSerArgArgSerSerProSerThrArgThrGly	6203
QY	488	GCAACTCGG---TGCGCCACAACTTTCGTCTCAAGACTGCTCGTCAAGTGTGCGCG	544
Db	6204	SerArgProAlaCysTrpThrAlaThrArg-----	6213
QY	545	ACCCTCGCGGCCCTGGGGCAAGACAACACTACTGGA-----TGCTCAACC	589
Db	6214	ThrProArgSerThrGlyProAlaSerThrSerGlyTrpArgProGlyProCysArgGly	6233
QY	590	CCAAACGCGAGTACACTTCG-----CCGACGGGCTTCGCGCGCGCGCGCGCA	637
Db	6234	ProSerValArgThrProAlaGlyArgLeuCysProValSerAlaThrAlaAlaProTrp	6253
QY	638	AGCGCCTCAGCACCGCGCGCGTCCCGCGCGCGGCTCGCGCCCGAGAGAGCCCGCG	697
Db	6254	ProMetTrpSerSerAsnArgArgProProArgProPro---AlaProArgArgSerArg	6272
QY	698	GCCTCCCGCGCGCCCGCGCGCGCGCGCGCG-----CCCGCGCCT	739
Db	6273	ProGluProCysSerArgCysProArgProProArgArgArgSerValThrAlaArgPro	6292
QY	740	CGC-----	742
Db	6293	ArgSerProSerGlySerArgArgAlaLeuThrTrpProArgSerGlyThrProTrpPro	6312
QY	742	-----	742
Db	6313	IleGlySerProArgTrpSerThrGlyArgProSerTrpArgProAlaGlyThrAsnTrp	6332
QY	743	---CCCGCATCGCTCCCGCCCGCCCGCCAGGAG-----AGCGGGCAGACCCCG	787
Db	6333	ProProGlyCysAlaArgTrpProLeuArgAsnProArgProValSerSerProGlyPro	6352
QY	788	CGGCAAGTTCTCCAGCT---CCTTCG-----CCATCG-----	817
Db	6353	ArgCysArgMetProAlaValProSerSerGlyCysSerArgValThrGlyProSerGlyPro	6372

Qy	818	-----ACAGCATCTCGCGAAGCCCTTCCGAGCCGTC-----	850
Db	6373	GlyTnpAlaalaAaCysTnpLysProSerArgSerSerProArgSerThrAsnTrpSer	6392
Qy	851	-----GCTCAGGAGACAGCGCCCGCGACGA-----	877
Db	6393	ArgCysSerArgArgSerGlySerArg---ProGlyArgCysCysTrpArgAlaThr	6411
Qy	878	---CGCTTCAGTGGGCGCGCCCTCCGCGCGCTGCCCG---	916
Db	6412	ThrProArgSerThrAlaProArgProSerSerArgCysSerGlySerProArgCys	6431
Qy	917	-----GCTTCCCCCG	925
Db	6432	GlyGlyHisGlyAlaSerSerProProSerSerAlaThrArgSerValArgSerPro	6451
Qy	926	CGCTCCTCCCG---CGCGCCCTCGCAGGCCCC---TGCTGC	961
Db	6452	ProProSerProProGlyArgProProThrGlyProGlySerAlaValGlyArgCysCys	6471
Qy	962	CGCTCTCGCGTACGGCGCGGGAGCGCGCGCGGTGG---1000	1000
Db	6472	---CysAlaArgProArgGlyValAlaArgTnpArgTnpSerValCysProSerThrArg	6490
Qy	1001	-----GGCGCGGAGCGCGGAGGTGCACGACGCGCGCCCTCTCGTTG	1048
Db	6491	ProProSerGlySerArgAlaMetThrArgTnpSerArgProSerArgPro-----6507	6507
Qy	1049	CACCTCTCCGGCGCGCGCCCGCCACGACATCCGAGCCCGCGCGG-----1096	1096
Db	6508	-----ArgArgArgProProSerProValThrProAlaArgSerArgArgSerSer	6524
Qy	1097	-----CGCGCG	1102
Db	6525	AlaAlaGlyProThrArgGlyTnpSerTyrGlyGlySerProProThrTnpProSerThr	6544
Qy	1103	CGCGGCACTGTACTGCCCTCG-----GGTGTGCCCGACGCGCTCGACGCGCGCTTAG	1156
Db	6545	AlaArgThrTnpThr---ProCysSerThrGlyCysAlaArgPro-----ProThr	6560
Qy	1157	TCCGNCCTCTGGCCCGCACCTCTGCTACCCGCTGGAGACGCTCC	1201
Db	6561	SerValProAlaHisArgThrArgArgSerThrArgArgArgSer	6575

## RESULT 14

```

RES001.14
US/10-084-846A-6
; Sequence 6, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNBERG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 6
; LENGTH: 19662
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid
; OTHER INFORMATION: Start codon: gga, Star
US/10-084-846A-6
Alignment Scores:

```

Pred. No.:	3.15e-10	Length:	19662
Score:	369.00	Matches:	182
Percent Similarity:	34.64%	Conservative:	21
Best Local Similarity:	31.06%	Mismatches:	166
Query Match:	16.43%	Indels:	217
DB:	15	Gaps:	35

US-10-087-080-31 (1-1209) x US-10-084-846A-6 (1-19662)

QY	1208	CAAGCTAGAGCGCTCTCCACCGGTTACAGACAGTGGCGGCCA---GGAGCGCGACTAAG	1152
Db	12641	ArgAlaProAlaProAlaGlyGlyGlyAlaGlyProTrpGlyArgValAlaSer	12660
QY	1151	CGCCCTGCAGGCTCGCGGCAGCCAGCGGGGCAGTCAGGTGGCGCGCCGCCGCCGCC	1092
Db	12661	ProAlaArgSerAlaArgAlaProAlaArgAlaProGlyAsnAlaAlaProProAlaThr	12680
QY	1091	GGGCTCGAGTGGCTTGGCGG---	1070
Db	12681	AspProArgThr-ThrProArgProArgProArgArgProAlaAlaArgCysProAl	12700
QY	1069	-----GGGCCGCCCGGAGAGGTGCAAGCAGGAGGGCGCGCGTTCGGTGGC	1020
Db	12700	aArgProAlaAlaArgProProGlySerAlaGlyArgAlaGlySerProProGlyValPr	12720
QY	1019	ACTCGGCCT-----	999
Db	12720	oProArgProProGlyArgAlaArgCysArgArgArgGlyAspArgAlaAlaGpArgPr	12740
QY	998	AGCGCGCGCG-----GCTCGCCGCGCGCTACGCGCAGA---	963
Db	12740	oSerProProProValAlaCysArgArgArgSerArgArgAlaArgProCysGlyAl	12760
QY	962	GGCAGCAGGCGCTTCGAGCGCCCGCGGGAGAGCGCGGGGAACG-----	909
Db	12760	aAlaGlyGlyProArgArgPro-----AlaArgSerThrSerArgArgProAl	12777
QY	908	GGCGGGC-----AGGGCGCGCGCCCTGGAAGCGTCTCCCGGGGCC	864
Db	12777	aAlaGlyProArgThrArgArgAlaAlaArgAlaProGluArgAlaAlaHisArgSp--	12796
QY	863	GTCTCCTGAGCGACGGCTGGGAAGGCTTGGCAGGATCTCTCAGATGGCGGAAGAG	804
Db	12797	-----ArgArgThrValAlaArgArgSerAsnAlaGsnArgSe	12809
QY	803	CTGGAGAACTTCCCGCGGGGTGGCGCCTCTCTCTGC---GGCGGGCGAGCGCATG	747
Db	12809	rArgValaArgProProArgProAlaAlaGlyArgProGlyArgGlyArgProAlaArgAl	12829
QY	746	CGGGCGAGCG	695
Db	12829	aGlyAlaArgProGlyProAlaAlaThrAlaProProAspArgArgGlyProAlaArgPr	12849
QY	694	-----GGGCT-----CCTCGGGCGCG	678
Db	12849	oSerGlyProArgHisGlyArgAlaArgArgGlySerGlyArgArgArgProArgProSe	12869
QY	677	AGCCCGGGCGCGGGACCGCGCGC---	653
Db	12869	rGlyArgProAlaAlaProProArgProArgGlyArgProGlySerAlaAlaAspProPr	12889
QY	652	-----GGTGGCTGAGCGCTTGGCGCGCGCGCGA---	612
Db	12889	oSerAlaProGlyGlySerSerSerCysProGlyArgGlyAlaLeuArgProArgAr	12909
QY	611	GGCA-----	608
Db	12909	gArgProProGlyThrProProAlaArgArgProGlyArgSerProArgArgGlyArgGl	12929
QY	607	-----AGGTGACTCGC-----TGTGGGGTTGAGCATCCAGTAGTGTG	570
Db	12929	yCysProGlyThrGlyGlyCysArgArgProAlaCysValaThrAlaAlaGlyArgCyl	12949

Qy	569	TCCT-----TGCCCGAGGCGCCG-----AGGGTCGCGCA--	539
Db	12949	sProAArgArgArgAlaCysProAlaAlaCysGlySerArgArgGlyArgAlaAr	12969
Qy	538	-----GACCTTGACGAGCAGTCGTTGACGAAAGTTCT-----	503
Db	12969	gProLeuAlaProArgArgTrpAsnAlaThrAlaGlyGlyAlaAlaArgGlyArgAlaAl	12989
Qy	502	-----GGCGCAGGAGTTGCGCCAGCCGCTGTAGTCCCGCGGAAAGGGGAC	453
Db	12989	aglyLeuArgGlySerArgSerAlaGlyAlaProThrSerProArgSerArgArgGlySe	13009
Qy	452	T-----TGCCCATGATGAGGTAC	438
Db	13009	rglyArgArgProArgArgArgGlyArgSerGlyProArgAlaGlyCysAlaValAlaSe	13029
Qy	437	T-----CGTTGATCTCCGCCAGCTCAAGCGCCGCGCGCGAGTCGCGGATGCGCCATGCGC	381
Db	13029	rValArgArgThrArgProProArgArgValArgProAlaSerArgGly-----	13045
Qy	380	ATGAGCGCATGTACAGTAGGGGGCTTGGCGCGCGGTATATGGCTTCTCGGTGCA	321
Db	13046	-----SerSerAlaArgProArgGlyAlaGlyThrAla-----	13056
Qy	320	CCCTCCCGCTCCCGC-----CGCGCGCGCGCGCTGGCGCGCGCGCCCGCCCTCCGCGGCC	264
Db	13057	-----ArgArgThrProGlyArgProArg-----ProProArgSerArgArgAr	13072
Qy	263	TCCGCCACACCGCTGCACAGCTGTCGCGGATCGCTCTCT-----	221
Db	13072	gAlaAlaProGly-----CysAlaGlySerProProSerThrArgArgArgAr	13088
Qy	220	-----CGCGCGCGCGCG-----CGCTCCCGCACACTGTGTCGCGCTGCCCTCGCTATCT	171
Db	13088	gSerProTrpProAlaGlyProArgAlaPro-----SerCysArgAlaAlaCysAlaValAr	13107
Qy	170	CTGG-----CGCGCGCGCGCGAGCGCTGG-----	143
Db	13107	gtTpAspSerAlaArgArgProArgAlaAlaGlyArgAlaGlyGlyProCysAlaGl	13127
Qy	142	-----CGCGCG-----AGTCCCCATCTGAGCCGAGG	117
Db	13127	yCysGlyArgProArgArgArgArgProAlaValArgProAlaProArgSerSerGl	13147
Qy	116	GAGTCGTCTCCGCGCGCGCACAGCGGG-----ACGGCGCGTCGCTG	75
Db	13147	yArgArgProSerArgProArgArgGlyHisArgArgSerArgTrpThrSerSerArgTh	13167
Qy	74	CGCG-----CGCGCGCTCCAGTCACTGCTGCTCCCGTGGCGCGCGCGGAGGAC	18
Db	13167	rArgTrpProArgProProArgHisCysProArgAlaAlaArgArgProArgArAl	13187
Qy	17	AACACCTCCCACTTCA	2
Db	13187	aaAppProSerProSer	13192
RESULT 15			
US-10-029-386-32327			
; Sequence 32327, Application US/10029386			
; Publication No. US20030194704A1			
; GENERAL INFORMATION:			
; APPLICANT: Penn, Sharron G.			
; APPLICANT: Rank, David R.			
; APPLICANT: Hanzel, David K.			
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL			
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO			
; FILE REFERENCE: AEOICA-X-2			
; CURRENT APPLICATION NUMBER: US/10/029,386			
; CURRENT FILING DATE: 2001-12-20			
; NUMBER OF SEQ ID NOS: 34286			
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1			
; SEQ ID NO 32327			
; LENGTH: 3237			

Search completed: March 29, 2004, 12:16:17  
Job time : 123 secs

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034344.19
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 6.8
; OTHER INFORMATION: SWISSPROT HIT: Q12949, EVALUATE 1.00e-102
; US-10-029-386-32327

```

<b>Alignment Scores:</b>					
Pred. No.:	8,05e-10	Length:	237		
Score:	366.50	Matches:	95		
Percent Similarity:	50.19%	Conservative:	32		
Best Local Similarity:	36.96%	Mismatches:	64		
Query Match:	16.74%	Indels:	66		
DB:	14	Gaps:	11		
US-10-087-080-31 (1-1209) x US-10-029-386-32327 (1-237)					
QY	214	GGCGGGGAGGCGATCCGCAGACGTCTCTCAGCGTGTGGCGGAGCGCGGAG	273		
Db	7	GlyGlyGlulnSerTyTrArgAlaAlaAlaAla-----AlaGlyGlyGlyTyR	24		
QY	274	GCGGGGCGCGCGGCCAGGCGG-----	297		
Db	25	ThrAlaMetProAlaProMetSerValTySerHisProIahisAlaGlulnTyPro	44		
QY	298	GGCGGCGCGGAGCGGAGGCGTCACGCAGCAAGCATATACGCGCGCGCCC-----	351		
Db	45	GlyGlyMetAlaArgAlaTyRGly-----ProTyThrProGlnProGlnPro	60		
QY	352	-----AAGCCCCCTACTCTGTACATCGCGCTCATCGGCATGGCCATCGCCGAC	399		
Db	61	LysAspMetVallysProProfySerTyRileAlaLeulleThrMetAlaIleGlnAsn	80		
QY	400	TGCGGGCGGCGGCTTGACGCTGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCC	459		
Db	81	AlaProAspLysIlelleThrLeuAsnGlylleTyRlInPheilleMetAspArgPhePro	100		
QY	460	TTTTTCGCGGAGCTACACAGGCTGGCGCAACTCTCGTGGCCACAACTTTTCGTCAAC	519		
Db	101	PheTyRArgAspAsnLysGlnGlyTrpGlnAsnSerileargHisAsnLeuSerLeuAsn	120		
QY	520	GACTCTTCGTCAAGTCTCGGACCCCTCGCGGCCCTGGGCAAGGAACAATACTG	579		
Db	121	GlucysPheValysValproArgaspAspylsysPro---GlyLysGlySerTyRTrp	139		
QY	580	ATGCTCAACCCCAACAGGAGTACACCTTCGCCACAGGGGTCTTCCGCCGCCGCCCAAG	639		
Db	140	ThrLeuAspProAspSerTyRAsnMetPheGluAnAnglySerPheLeuLeuAspArgA	159		
QY	640	CGCCTCAGC-----CACCGCGCG	657		
Db	160	ArgPheLysLysLysAspAlaVallysAspLysGluGluLysAspArgLeuHis-Leu	179		
QY	658	CGGTCGCCGCGCGCGGCTGGGCGCCAGAGAGGCCCGGGCTTCCGCCGCCGCCCGCG	717		
Db	179	sGluProProProProGly-----ArgGlnProProProAlaProGluGlnAl	196		
QY	718	CCGCGCGCGCGCGCGGCTCGCCCGCATGCGCTCGCCCGCGCGCGCAGG-----	769		
Db	196	aAspLyAsnAlaProGlyProGlnPro-----ProProValArgIleGlnAs	212		
QY	770	-----AGGAGCGCGCCAGCCCGCGGCAAGTTCTCCAGCTCCTTCGCC	813		
Db	212	pIleLysThrGluAsnGlyThrCyseProSerPro-----ProGlnPr	226		
QY	814	ATCAGACAGCATCTCGCGAAGCCCTTCGCGACGCCCTCGCTCAGGACA	862		
Db	226	o-----LeuSerProAlaAlaAlaLeuGlySerGlySer	237		

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 29, 2004, 11:57:01 ; Search time 26.5 Seconds  
(without alignments)  
8777.023 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2189

Sequence: 1 atgaagtggagggtgttctg.....tggagagctctagtgtga 1209

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=cpn2.1/USPTO spool.p/US10087080/runat\_29032004\_114527\_6563/app\_query.fasta\_1.1351  
-DB=pir\_78 -QFMT=fasta -SUFFIX=n2p.rpr -MINMATCH=0.1 -LQOPEXT=0  
-UNITS=BITS -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10087080 @CCN 1.1.38 @runat\_29032004\_114527\_6563 -NCPU=3  
-NO MAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1314.5	60.1	439	2 I60916	HNF-3/forkhead hom
2	543.5	24.8	440	2 S71795	transcription fact
3	517	23.6	465	2 G02738	PREAC-4 - human
4	431.5	19.7	444	2 T09474	forkhead protein F
5	387	17.7	473	2 S70357	forkhead transcrip
6	380.5	17.4	376	2 I49735	hepatocyte nuclear
7	380	17.4	461	2 S34472	MRF-1 protein mo
8	366.5	16.7	468	1 A54258	transcription fact
9	358	16.4	476	2 A54743	transcription fact
10	356	16.3	459	1 B54258	transcription fact
11	356	16.3	466	2 A36674	transcription fact
12	348.5	15.9	354	1 S35090	transcription fact
13	345	15.8	480	2 JH0672	brain factor 1 pro
14	344.5	15.7	458	2 B39533	transcription fact

15 344 15.7 367 2 S49009 fork head protein  
16 343.5 15.7 469 2 I37451 HBF-02 (HFK-2) pro  
17 338.5 15.5 370 2 S49008 fork head protein  
18 338 15.4 333 2 T20033 hypothetical prote  
19 335.5 15.3 353 1 C54258 transcription fact  
20 327 14.9 106 2 S51624 FREAC-1 protein - Af  
21 326.5 14.9 427 2 I51580 XfKH2 protein - Af  
22 326 14.9 106 2 S51625 FREAC-2 protein - Af  
23 324 14.8 508 2 S59870 fork head domain p  
24 324 14.8 510 2 A32380 nuclear protein f  
25 323 14.8 680 1 Q0B230 BHLFI protein - hu  
26 320.5 14.6 106 2 S51627 forkhead protein F  
27 320 14.6 451 2 A55909 transforming prote  
28 319.5 14.6 128 2 C46178 probable transcrip  
29 318.5 14.6 839 2 F75518 hypothetical prote  
30 315 14.4 349 2 A56365 silk gland factor-  
31 315 14.4 445 2 S23055 slp2 protein - fru  
32 313.5 14.3 101 2 I60922 brain factor-2 - r  
33 313.5 14.3 322 2 S23053 sloppy paired prot  
34 313 14.3 445 2 S23056 slp2 protein - fru  
35 312 14.3 1585 2 T31611 hypothetical prote  
36 311.5 14.2 101 2 I60919 HNF-3/fork-head ho  
37 311.5 14.2 387 2 A47446 HNF-3/fork head fa  
38 308.5 14.1 111 2 C47746 fork head homolog  
39 308.5 14.1 111 2 G56556 HNF-3/fork-head ho  
40 303.5 13.9 101 2 I60917 HNF-3/fork-head ho  
41 302.5 13.8 101 2 A47450 transcription fact  
42 302.5 13.8 534 2 S71800 HNF-3/fork-head ho  
43 301.5 13.8 101 2 I60920 transcription fact  
44 301.5 13.8 111 2 B47746 forkhead transcrip  
45 299.5 13.7 434 2 I51436 HNF-3beta - Africa

#### ALIGNMENTS

##### RESULT 1

I60916  
HNF-3/forkhead homolog-1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 02-Jul-1996 #sequence, revision 02-Jul-1996 #text\_change 24-Sep-1999  
C:Accession: I60916  
R:Clcvidence, D.E.; Overdier, D.G.; Tao, W.; Qian, X.; Pani, L.; Lai, E.; Costa, R.H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 3948-3952, 1993  
A:Title: Identification of nine tissue-specific transcription factors of the hepatocyte  
A:Reference number: A47450; MUID:93248207; PMID:7683413  
A:Accession: I60916  
A>Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: mRNA  
A:Residues: 1-439 <RES>  
A:Cross-references: GB:L13201; NID:g951067; PIDN:AAA74561.1; PID:g550513  
C:Genetics:  
A:Gene: HNF-1  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F100-192/Domain: fork head DNA-binding domain homology <PHD>

Alignment Scores:  
Pred. No.: 4.42e-58 Length: 439  
Score: 1314.50 Matches: 278  
Percent Similarity: 74.30% Conservative: 14  
Best Local Similarity: 70.74% Mismatches: 86  
Query Match: 60.05% Indels: 15  
DB: 2 Gaps: 6

US-10-087-080-31 (1-1209) x I60916 (1-439)

QY 49 GGCGAGTCCTGGAGGGCGGGCGGCGAGCGCGCTCCCGCTGTGGCGCGCGGA 108  
Db 2 GlySerAspLeuGluGlyAlaGlySerAspValProSerProLeuSerAlaAlaGly 21  
QY 109 GACGACTCCCTGGCTCAGATGGGAGCTGCGCGCC---AAGCGTCCGCGGCGCGGC 165  
Db 22 AspAspSerLeuGlySerAspGlyAspCysAlaAlaAsnSerProAlaAlaGlySerGly 41



```
Db      270  lNProProCysAlaTyTyGlyProTyTyGlyCysGlyTyTyGlyLeuGlnLeuGlnProTyTyH 290
Qy      765  CCAGGAGGAGCGG-----CGAGCCCGGGGCAAGTTCTCCAGTTC 806
Db      290  iProHisSerAlaLeuPheAlaPheHisHisProSerProProProAlaArgGlnProProA 310
Qy      807  CTTGCCCATCGACAGCATCTCGCGAAGCCCTTCGCGACCGCTCGCTCAGGACACGCG 866
Db      310  laAlaProAlaGlyAla-ProAlaAlaLeuProProProProProProProProPro 329
Qy      867  CCGCGGAGCAGCGCTTCTAGTGGGGCGCCGCGC-----CTCCCGCGCGCT-----GCC 914
Db      330  ProArg-----ArgArgAlaProLeuLeuProAlaAlaGluLeuAla 343
Qy      915  CGCGTTCCCGCGCTCTCCCGCGCGCCCTGCAGGCGCCCTGCGCGCTCTGCGCGTA 974
Db      344  ArgThProPheGlyTyTyProHisProLeu-----GlyProAlaLeuAlaAlaSerLeu 361
Qy      975  CGCGCGCGCGAGCGCGCGCGTGGCGCGCGCGAGCGCGAGGTGCCACCGCGCC 1034
Db      362  HisAlaAlaLysProGlySer-----GlyAlaAlaValAlaArg 374
Qy      1035  GCCCTCTCTGTTGCACTCTCT-----CCGCGCGCGCGCCCGCCAA-----GCCACT 1082
Db      375  SerProPheSerIleGluSerIleIleGlyGlyProGlyProGlyLeuGlyAlaGly 394
Qy      1083  CCGAGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1133
Db      395  ProAlaProGlyAlaGlyGlySerCysAlaSerGlnSerGlyAlaAlaThrGlyLeuSer 414
Qy      1134  CGAGCGCTGCGAGCGCGCTTGTAGTCGNGCTGCTGCGCGCGCGCTGCTGCTGCTGCTGCT 1190
Db      415  ArgSerLeuGlySerGlyLeuAlaProAlaAla---AlaLeuProAlaAlaProGly 432

RESULT 3
G02738
FREC-4 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 19-May-2000
C:Accession: G02738
R:Enrback, S.
submitted to the EMBL Data Library, June 1996
A:Reference number: H01646
A:Accession: G02738
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-465 <ENE>
A:Cross-references: EMBL:U59832; NID:g1399238; PID:g1399239
C:Genetic:
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F:125-216/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:
Pred. No.:      1,12e-18      Length:      465
Score:          517.00      Matches:     157
Percent Similarity: 44.21%      Conservative: 34
Best Local Similarity: 36.34%      Mismatches: 132
Query Match:      23.62%      Indels:     109
DB:              2          Gaps:         16

US-10-087-080-31 (1-1209) x G02738 (1-465)

Qy      60  GGAGGCGGCGGCGGCGAGCGCGCGTCCCGTCTGCGGCGGCGAGCGACTCCCT 119
Db      39  GlyGlyGlyGlyGlyProArgLeuAlaValPro----- 49
Qy      120  GGCGTCCAGATGGGAGCTGCGCGCCCAAGCGCTCCGCGGCGCGCGCGAGATACGA 179
Db      50  -----AlaGlnArgArgArgArgSerTyAla 60
Qy      180  GGGCGAGCGGCAACAGAG-----TGGCGG 203
```

```
Db      61  GlyGluAspGluLeuGluAspLeuGluGluGluAspAspAspAspIleLeuLeuAla 80
Qy      204  AGGCGGCGCGGCGCGGAGGAGGAGGATCCCGGAGAGAGCTGCTGCACGCGTGTGGCGGA 263
Db      81  ProProAlaGlyGlySerProAlaProProGlyProAlaProAlaAlaGlyAlaGlyAla 100
Qy      264  GGGCGCGGAGCGCGGCGCGGCGCGGCGCGGCGCGGCGGAGCGCGCGCGGCGTGC 323
Db      101  GlyGlyGlyGlyGlyGlyGlyAla-GlyGlyGlyGlySer---AlaGlySerGlyAl 119
Qy      324  ACGCAGCAGCATATACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 383
Db      119  aLysAsn---ProLeuVal-----LysProProTyTySerTyTyIleAlaLeuIleTh 135
Qy      384  CATCGCCATCCGACCTCCGCGCGCGCGCGCTTGACGCTGCGGAGAGATCAACAGATACCT 443
Db      135  rMetAlaIleLeuGlnSerProLysLysArgLeuThrLeuSerGluIleCysGluPheI 155
Qy      444  CATGGGCAAGTTCCTTTTTCGCGCGCGAGCTACACGGGCTGGCGCAACTCCGTGGCGCA 503
Db      155  eSerGlyArgPheProTyTyArgGluLysPheProAlaTrpGlnAsnSerIleArgH 175
Qy      504  CAACCTTTCTCTCAACGACTGCTTCAAGGTGCTGCGGAGCGCGCGCGCGCGCGCGCG 563
Db      175  sAsnLeuSerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro---G 194
Qy      564  CAAGGACAACCTACTGTGATGCTCAACCCCAACAGCAGAGTACACCTTCGCCGCGGGTCT 623
Db      194  YLysGlyAsnTyTyTrpThrLeuAspProGluSerAlaAspMetPheAspAsnGlySerP 214
Qy      624  CCGCGCGCGCGCGAGCGCTCAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
Db      214  eLeuArgArgLysArgPheLys---ArgGlnProLeuLeuProProAsnAlaAla 233
Qy      684  CGAGGAGCGCC-----CCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 710
Db      233  aAlaGluSerLeuLeuArgGlyAlaGlyAlaAlaGlyGlyAlaGlyAspProAlaAl 253
Qy      711  CCGG-----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 742
Db      253  aAlaAlaAlaLeuPheProProAla-ProProProProHisAlaTyTyGlyTyGlyP 273
Qy      743  -----CCGCGCATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
Db      273  roTyTyGlyCysGlyTyTyGlyLeuGlnLeuProProTyTyAlaProProSerAlaLeuPheA 293
Qy      774  GCGCGCGCGCGCGCGCGGCAAGTTCCTCCAGCTCTTCGCCATCGACAGCATCTGCCCAA 833
Db      293  laAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaPheHisProHis-----S 308
Qy      834  GCCCTTCGCGAGCGCTCGCTCAGGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 893
Db      308  erPro-ProProProProPro---HisGly----- 317
Qy      894  CGCGCGCTCCCGCGCGCTGCGCGCGGTTCGCCCGCGCT-----CCT 932
Db      318  -----AlaAlaAlaGluLeuAlaAlaArgThrAlaPheGlyTyTyArgProHisPro 333
Qy      933  CCGCGCGCGCGCGCGCGCGCGCTGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 992
Db      334  LeuGlyAlaAlaLeuProGlyProLeuProAlaSerAlaAlaLysAlaGlyGlyProGly 353
Qy      993  CGCGCTCGGCGCGCGCGAGGC-----CGAGTGCACCGC 1028
Db      354  AlaSerAlaLeuAlaArgSerProPheSerIleGluSerIleIleGlyGlySerLeuGly 373
Qy      1029  CGCGCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1088
Db      374  ProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 393
Qy      1089  CCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1148
Db      394  Pro-----SerProValAlaAlaProProAlaProGlySerSerGlyGly 408
```

```
QY 1149 GGCTTAGTCGCGTCCTGGCCGCGACCT 1178
    |||
    |||
    |||
Db 409 GlyCysAlaAlaGlnAlaAlaValGlyPro 418
    |||
    |||
    |||
RESULT 4
T09474
forkhead protein FRAC-2 - human
C;Species: Homo sapiens (man)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C;Accession: T09474
R;Helqvist, M.; Mahlapuu, M.; Blixt, C.; Enerback, S.; Carlsson, P.
J. Biol. Chem. 273, 23335-23343, 1998
A;Title: The human forkhead protein FRAC-2 contains two functionally redundant activation domains
A;Reference number: Z15682; MUID:98399768; PMID:9722567
A;Accession: T09474
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-444 <HEL>
A;Cross-references: EMBL:U13220; NID:G3425849; PID:G3425850
A;Experimental source: lung
C;Function:
A;Description: may function as transcription activator
C;Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
F;100-191/Domain: fork head DNA-binding domain homology <FHD>
Alignment Scores:
Pred. No.: 1.9e-14 Length: 444
Score: 431.50 Matches: 129
Percent Similarity: 42.82% Conservative: 41
Best Local Similarity: 32.49% Mismatches: 156
Query Match: 19.71% Indels: 69
DB: 2 Gaps: 11
US-10-087-080-31 (1-1209) x T09474 (1-444)
QY 73 GCAGAGCGCGGCGTCCCGTCTCGCGGGGAGAGACTCCTCGGTCTCA---GAT 129
    |||
    |||
    |||
Db 5 GlyGlyProProAlaProLeuArgAlaCysSerProValProGlyAlaLeuGln 24
    |||
    |||
    |||
QY 130 GGGGACTCGCGCCCAAGCGTCCGCGGGCGGGCGCCAGAGATACGAGCGGCGCGG 189
    |||
    |||
    |||
Db 25 AlaAlaLeuMetSerProProAlaAlaAlaAlaAlaAlaAlaProGluThr 44
    |||
    |||
    |||
QY 190 GAACAGAGTCGGAGCGGGCGCGGCGGAGGAGCGATCCGCGCACACTGCTGCA 249
    |||
    |||
    |||
Db 45 ThrSerSerSerSerSerSerAlaSerCysAlaSerSerSerSerSerSer 64
    |||
    |||
    |||
QY 250 GCGGTGTCGCGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 309
    |||
    |||
    |||
Db 65 SerAlaSerAlaProSerAlaAlaCysLysSerAlaGlyGlyAlaGlyAlaGly 84
    |||
    |||
    |||
QY 310 AGCGGCGAGGTCGCGCAGCAAGCATATACGCGCGGCGCC---AAGCCCTTACTCG 366
    |||
    |||
    |||
Db 85 SerGlyGlyAlaLysLysAlaSerSerGlyLeuArgArgProGluLysProProTyrSer 104
    |||
    |||
    |||
QY 367 TACATCGCTCATCTCCATGCGCATCCGCGGCGGCGGCGGCGGCGGCGGCGG 426
    |||
    |||
    |||
Db 105 TyrAlaLeuLeuValMetAlaIleGlnSerSerSerLysArgLeuThrLeuSer 124
    |||
    |||
    |||
QY 427 GAGATCAACGAGTACCTCATGCGCAAGTCCCTTTTCGCGGCGAGTACACGCGCTGG 486
    |||
    |||
    |||
Db 125 GluIleTyrGlnPheLeuGlnAlaArgPheProPheArgGlyAlaTyrGlnGlyTrp 144
    |||
    |||
    |||
QY 487 CGAATCGCTGCGGCACACCTTTCGTCACAGACTGTTGTCAGGTGTCGCGGAC 546
    |||
    |||
    |||
Db 145 LysAsnSerValArgHisLeuSerLeuAsnGluCysPheIleLysLeuProLysGly 164
    |||
    |||
    |||
QY 547 CCCTCGCGGCGGCGGCGAGGACAACTACTGTGATGCTCAACCCCAACAGCGAGTACAC 606
    |||
    |||
    |||
Db 165 LeuGlyArgPro---GlyLysGlyHisTyrTrpThrIleAspProAlaSerGluPheMet 183
    |||
    |||
    |||
QY 607 TTGCGCGAGCGGGTCTTCGCGCGCGCGCGAGCGCTCAGCACCGCGCGGCTCC 666
    |||
    |||
    |||
Db 184 PheGluGluGlySerPheArgArgArgProArgGlyPheArgArgLysCysGlnAlaLeu 203
    |||
    |||
    |||
QY 667 GCGCCC-----GGGCTG-----
    |||
    |||
    |||
Db 204 LysProMetTyrHisArgValValSerGlyLeuGlyPheGlyAlaSerLeuLeuProGln 223
    |||
    |||
    |||
QY 679 -----CGGCGCGAGGAGGCGCCG-----
    |||
    |||
    |||
Db 224 GlyPheAspPheGlnAlaProProSerAlaProLeuGlyCysHisSerGlnGlyTyr 243
    |||
    |||
    |||
QY 697 ---GGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
    |||
    |||
    |||
Db 244 GlyGlyLeuAspMetMetProAlaGlyTyrAspAlaGlyAlaGlyAlaProSerHisAl 263
    |||
    |||
    |||
QY 754 TCGCGCGCGCGCGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
    |||
    |||
    |||
Db 263 AHISPROHISHISHISHISHISHISHISHISHISHISHISHISHISHISHISHISH 283
    |||
    |||
    |||
QY 814 ATCAGACAGCATCTCTGCGCAAGCCCTTCGCGAGCGCGTCCGCTCAGGGACACGCGCGCGG 873
    |||
    |||
    |||
Db 283 rTyrMetAlaSerCysProValProAlaGly-----ProGlyG 296
    |||
    |||
    |||
QY 874 ACGAGCTTACGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 933
    |||
    |||
    |||
Db 296 yValGlyAlaAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 314
    |||
    |||
    |||
QY 934 CCGCGCGCGCGCTGCGAGGCGCGCTGCTGCGCGCTGCTGCGCGCTGCTGCGCGCGCGCG 993
    |||
    |||
    |||
Db 314 rSerProValProSerSerSerProAlaMetAlaSerAla----- 326
    |||
    |||
    |||
QY 994 CGGTGGCGCGCGCGAGGCGCGGAGTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 1049
    |||
    |||
    |||
Db 327 -----IleGluCysHisSerPro-TyrThrSerProAlaAlaHisT 340
    |||
    |||
    |||
QY 1050 --ACTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1107
    |||
    |||
    |||
Db 340 rPsrSerProGlyAlaSerProTyrLeuLysGlnPro----- 352
    |||
    |||
    |||
QY 1108 CACTGTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 1154
    |||
    |||
    |||
Db 353 --ProAlaLeuThrProSerSerAsnProAlaAlaSerAlaGlyLeu 367
    |||
    |||
    |||
RESULT 5
S70357
forkhead transcription factor HNF-3 alpha - human
N;Alternate names: hepatocyte nuclear factor-3 alpha
C;Species: Homo sapiens (man)
C;Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 16-Jul-1999
C;Accession: S70357
R;Bingle, C.D.; Gowan, S.
Biochim. Biophys. Acta 1307, 17-20, 1996
A;Title: Molecular cloning of the forkhead transcription factor HNF-3-alpha from a human
A;Reference number: S70357; MUID:96254057; PMID:8652662
A;Accession: S70357
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-473 <BIN>
A;Cross-references: EMBL:U39840; NID:gl066121; PID:AB06493.1; PID:gl066122
C;Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
F;171-262/Domain: fork head DNA-binding domain homology <FHD>
Alignment Scores:
Pred. No.: 2.97e-12 Length: 473
Score: 387.00 Matches: 124
Percent Similarity: 44.36% Conservative: 45
Best Local Similarity: 32.55% Mismatches: 132
Query Match: 17.68% Indels: 80
DB: 2 Gaps: 15
US-10-087-080-31 (1-1209) x S70357 (1-473)
QY 70 GCGCGAGCGAGCGCGCGTCCCGCTGTCGCGGGGAGAGAGACTCCTCGGTGGCTCAGAT 129
    |||
    |||
    |||
```

```

Db      88 GlySerAlaGlyAlaMetAsnSerMetThrAlaAlaGlyValThrAlaMetGlyThr--- 106
QY      130 GGGGACTGCGCGCAAGCGTCCGCGCGCGCGCGCCAGAGATACGACGAGCGGACGCGC 189
Db      107 -----AlaLeuSerProSer---GlyMetGlyAlaMetGlyAlaGlnAlaAla 122
QY      190 GAACAGAGTGGGAGGCGGCGCGCGCGGAGGAGGAGGATCCCGGACAGCTGCTGCA 249
Db      123 SerMetMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCysMetSerProMet 142
QY      250 GCGGTGGTGGCGAGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 309
Db      143 AlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGly-----GlyGlyGlyAsp 159
QY      310 AGCGCGAGGTGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 369
Db      160 AlaLysThrPheLysArgSerTyrProHisAla-----LysProProTyrSerTyr 176
QY      370 ATCCGCGCTCATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCG 429
Db      177 IleSerLeuIleThrMetAlaIleGlnArgAlaProSerLysMetLeuThrLeuSerGlu 196
QY      430 ATCAACAGTACTCATGCGCAAGTTCCTTTTCCGCGCGAGCTACACGGGTGCGGC 489
Db      197 IleTyrGlnTrpIleMetAspLeuPheProTyrTyrArgGlnAsnGlnArgTrpGln 216
QY      490 AACTCCGTGGCGCAACCTTCGTCAACGACTGCTTCGTCAAGTCTCTCGCGAGCCCGC 549
Db      217 AsnSerIleArgHisSerLeuSerPheAsnAspCysPheValLysValAlaArgSerPro 236
QY      550 TCGCGCGCTCGGCGCAAGCAACTACTGATGTCTCAACCCCAACAGCAGTACACCTTC 609
Db      237 AspLysPro---GlyLysGlySerTyrTrpThrLeuHisProAspSerGlyAsnMetPhe 255
QY      610 GCCGACGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
Db      256 GluAsnGlyCysTyrLeuArgArgGlnLysArg----- 266
QY      670 CCGCGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
Db      267 -----PheLysCysGluLysGlnProGly-----AlaGlyGly 277
QY      730 GCCCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783
Db      278 GlyGlyLysGlySerGlyLysGlySerGlyAlaLysGlyProGluSerArgLysAsp 297
QY      784 CCGCGGCAAGTCTCAGCTCCTCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGC 843
Db      298 ProSerGly-----AlaSerAsnProSerAlaAspSerProLeuHisArgGlyValHis 315
QY      844 AGCGTGCCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
Db      316 -----GlyLysThrGlyGlnLeuGluGlyAlaProAla 326
QY      904 CCGCGTGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 963
Db      327 ProGlyProAlaAlaSerProGlnThrLeuAsp----- 337
QY      964 CTCTGCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1023
Db      338 -----HisSerGlyAlaThrAlaThrGlyGlyAlaSerGluLeuLysThrPro 353
QY      1024 -----CCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
Db      354 AlaSerSerThrAlaProProIleSerSerGlyProGlyAlaLeuAlaSerValProAla 373
QY      1063 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
Db      374 SerHisProAlaHisGlyLeu---AlaProHisGluSerGlnLeuHisLeu----- 389
QY      1123 CTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182

```

```

Db      390 -----LysGlyAspProHisTyrSer 396
QY      1183 TAC 1185
Db      397 Phe 397
RESULT 6
I49735
hepatocyte nuclear factor 3 forkhead homolog 8 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 24-Sep-1999
C/Accession: I49735
R/Clevidence: D.E.; Overdier, D.G.; Peterson, R.S.; Porcella, A.; Ye, H.; Paulson, K.E.;
Dev Biol 166, 195-209, 1994
A/Title: Members of the HNF-3/forkhead family of transcription factors exhibit distinct
A/Reference number: I49735; MUID:95046902; PMID:7958446
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-376 <RES>
A/Cross-references: GB:I35949; NID:G762833; PIDN:AAA64885.1; PID:G575512
C/Genetics:
A/Gene: HNF-8
C/Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F;47-138/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:
Pred. No.: 6,48e-12 Length: 376
Score: 380.50 Matches: 112
Percent Similarity: 44.65% Conservative: 34
Best Local Similarity: 34.25% Mismatches: 100
Query Match: 17.38% Indels: 81
DB: 2 Gaps: 12

US-10-087-080-31 (1-1209) x I49735 (1-376)
QY      219 GGAGGAGGAGTATCCCGGAGAGTCTCTCAGCGGTGTGGCGGAGGCGCGGAGCGCG 278
Db      12 GlyGlyGly-----ThrGlyGlyGlyGlyGlyGlyGln 24
QY      279 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338
Db      25 -AlaMetAspProAlaAlaAlaProHisGlnGlyGlnGluAspGlnArgArg----- 42
QY      339 TAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398
Db      43 -AlaArgProGluLysProProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSe 62
QY      399 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 458
Db      62 rSerProSerLysArgLeuThrLeuSerGluIleTyrGlnPheLeuGlnAlaArgPhePr 82
QY      459 CTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 518
Db      82 OphePheArgGlyAlaTyrGlnGlyTrpLysAsnSerValArgHisAsnLeuSerLeuAs 102
QY      519 CGACTGCTTCGTCAAGTGTCTGCGGAGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 578
Db      102 nGluCysPheIleLysLeuProLysGlyLeuGlyArgPro---GlyLysGlyHisTyrTr 121
QY      579 GATGTCACACCCCAACAGCAGTACACCTTCGCCACGCGGTCTTCGCCCGCGCGCGCGCA 638
Db      121 pThrIleAspProAlaSerGluPheMetPheGluGluGlySerPheArgArgProAr 141
QY      639 GCGCTCAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680
Db      141 gglyPheArgArgLysCysGlnAlaLeuLysProValTyrSerMetValAsnGlyLeuGl 161
QY      681 GCCCGAGGAGGCGCGCG-----GCCCTCCCGCGCGCGCGCGCGCGCGCGCGCG 713
Db      161 yPheAsnHisLeuProAspThrTyrGlyPheGlnGlySerGlyGlyLeuSerCysAlaPr 181
QY      714 GCCG----- 717

```





```

QY 82 GCGCGTCCCGCTGTGCGCGCGGAGACGACTCCCTGGCTCAGATGGGAGTGGCG 141
Db 77 ProProProLeuAlaA-----
QY 142 GCCAAGCGCTCGCGCGCGCGGCGCAGAGATACGCGAGCGCGACGCGCAGACAGATGGC 201
Db 83 -----ProGlnAlaGlyGlyAlaAAGlnSerAsnAspGluLysGlyProGlnLeuLeu 100
QY 202 GGAGCGCGCGCGCGGAGGAGCGATCCCGCAGCAGCTGCTCGACGGGTGGCG 261
Db 101 LeuLeuProProThrAspHisHisArgProProSerGlyAlaLysAlaGlyGlyCysCys 120
QY 262 GAGCGCGCGAGCGCGCGCGCGCGCA-----GCGCGCGCG 300
Db 121 ArgProGlyGluLeuGlyProValGlyProAspGluLysGlyAlaGlyAlaGly 140
QY 301 GGC-----GCGGAGCGCGGAGGTGCACG----- 327
Db 141 GlyGluGluLysGlyAlaGlyGluGlyGlyLysAspGlyGluGlyGlyLysGluGly 160
QY 328 AGCAAGCATATACGCGCGCGCGCAAGCCCTACTCGTACATCGCGCTCATCGCATG 387
Db 161 GluLysLysAsnGlyLysTyrgluLysProPheSerTyrglnAlaLeuLeuMetMet 180
QY 388 GCCATCCGCGACTCGCGCGCGCGCTTGACGCTGCGGAGATCAACGAGTACTCTCATG 447
Db 181 AlaIleArgGlnSerProGluLysArgLeuThrLeuAsnGlyIleTyrgluPheLeuMet 200
QY 448 GGCAAGTTCCTCTTTCGCGCGCGCTACACGGCTGGCGCAACTCCGTGGCGCCACAC 507
Db 201 LysAsnPheProTyrglyTyrgluLysGlnGlyTyrglnAsnSerIleArgHisAsn 220
QY 508 CTTTCGCTCAACGACTCTGCTCAAGTGTCTGCGGACCCCTCGCGGCGCTCGGCAAG 567
Db 221 LeuSerLeuAsnLysCysPheValLysValProArgHisTyrglnAspPro----GlyLys 239
QY 568 GACAATCTAGTGTCTCAACCCCAACAGAGTACACCTTCGCCGCGGCTTCGCG 627
Db 240 GlyAsnTyrglnMetLeuAspProSerAspValPheIleGlyThrThrGly 259
QY 628 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 260 LysLeuArgArgSerThrThrSerProAlaLysLeuAlaPheLysArgGlyAlaAla 279
QY 661 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 702
Db 280 LeuThrSerThrGlyLeuThrPheMetAspAlaProAlaProSerThrGlyProCysArg 299
QY 703 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732
Db 300 ProSerCysProCysThrThrProAlaSerSerThrLeuSerTyrglnGlyThrSer 319
QY 733 -----CGGCGCTGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 786
Db 320 AlaTyrglnProSerHisProMetProTyrglnSerSerValLeuThrGlnAsnSerLeuGlyAsn 339
QY 787 GCGGCAAGTTCCTCACTCC-----TTGCGCATGACAGCATCTCTGCGCAAG----- 834
Db 340 AsnHisSerPheSerThrAlaAsnGlyLeuSerValAspArgLeuValAsnGlyGluLeu 359
QY 835 CCTTCGCGAGCGCTGCGCTCAGGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 891
Db 360 ProTyrglnAlaHisLeuThrAlaAlaAlaLeuAlaAlaSerValProCysGlyLeu 379
QY 892 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
Db 380 SerValProCys-----SerGlyThrTyrglnSerLeuAsnProCysSer 393
QY 952 GCCTCTGCTG-----CCGCTCTGCGGTAC 975
Db 394 ValAsnLeuLeuAlaGlyGlnThrSerTyrglnPheProHisValProHisProSerMet 413
QY 976 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1035

```

```

Db 414 ThrSerGlnSerSerThrSerMetSerAlaArgAlaAlaSer-----SerSerThrSerPro 432
QY 1036 CCCTCTCTCTCTGACCTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1086
Db 433 Pro-----AlaProArgPro-----LeuProCysGluSerLeuArg 444

RESULT 10
B54258
transcription factor HNF-3 beta - mouse
N:Alternate names: hepatocyte nuclear factor 3 beta
C:Species: Mus musculus (house mouse)
C:Date: 18-Aug-1995 #sequence revision 19-Oct-1995 #text_change 18-Jun-1999
C:Accession: B54258; I49676; S37185
R:Kaestner, K. H.; Hiemisch, H.; Luckow, B.; Schutz, G.
Genomics 20, 377-385, 1994
A:Title: The HNF-3 gene family of transcription factors in mice: gene structure, cDNA se
A:Reference number: A54258; MUID:94307723; PMID:8034310
A:Accession: B54258
A:Molecule type: mRNA
A:Residues: 1-459 <KAE>
A:Cross-references: EMBL:X74937; NID:9402190; PIDN:CAA52891.1; PID:9402191
R:Sasaki, H.; Hogan, B. L.
Development 118, 47-59, 1993
A:Title: Differential expression of multiple fork head related genes during gastrulation
A:Reference number: I49674; MUID:93387221; PMID:8375339
A:Accession: I49676
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-9, 'H', 11-459 <RES>
A:Cross-references: GB:L10409; NID:9404763; PIDN:AAA03161.1; PID:9404764
C:Genetics:
A:Gene: HNF-3beta
A:Introns: 23/3
C:Function:
A:Description: regulation of gene expression, especially in liver and lung
A>Note: expressed in embryonic endoderm and endoderm-derived adult tissues, most strongly
C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
C:Keywords: DNA binding; transcription factor
F:159-250/Domain: fork head DNA-binding domain homology <PHD>

Alignment Scores:
Pred. No.: Length: 459
Score: 356.00
Matches: 120
Percent Similarity: 41.49%
Conservative: 36
Best Local Similarity: 31.91%
Mismatch: 148
Query Match: 16.26%
Indels: 72
DB: 1
Gaps: 11

US-10-087-080-31 (1-1209) x B54258 (1-459)
QY 49 GCGAGTACCTGGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGA 108
Db 54 GlySerGlyAsnMetSerAlaGlySerMetAsnMetSerSerTyrglnGlyAlaGlyMet 73
QY 109 GACGACTCCCTGGGCTCAGATGGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 168
Db 74 SerProSerLeu-----AlaGlyMetSerProGlyAlaGlyAlaMetAla 88
QY 159 AGAGATACGAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 225
Db 89 GlyMetSerGlySerAlaGlyAlaGlyValAlaGlyMetGlyProHisLeuSerPro 108
QY 226 CGCATC-----CCGCGCAGCAGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 270
Db 109 SerLeuSerProLeuGlyGlyGlnAlaAlaGlyAlaMetGlyGlyLeuAlaProTyrgln 128
QY 271 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 330
Db 129 AsnMetAsnSerMetSerProMetTyrglnAlaGlyLeuSerArgAlaArgAspPro 148
QY 331 AAGCATATACGCGG-----CGGCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTATC 381

```



Db 374 AlaProHisGluSerGlnLeuHisLeuLysGlyAlaProHisTySerPheAsnHisPro 393  
 QY 808 TTCGCATCGACAGCATCTCGCGAAGCCTTCGCGACCGCTCGCTCAGGACACGCGCC 867  
 Db 394 PheSerIleAsnAsnLeuMetSerSerGluGlnGlnHisLeuAspPheLysAla 413  
 QY 868 CCGCGACGACGCTTCAGTGGGGCGCGCGCTCGCGCTCGCGCTCGCGCTCGCGCT 927  
 Db 414 TyrGluGlnAlaLeuGlnTySer 421  
 QY 928 CTCCTCCCGCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCG 987  
 Db 422 422  
 QY 988 CCGCGCGCTCG 1041  
 Db 428 ProAlaSerLeuProLeuGlyGlyAlaSerValAlaThrArgSerProIleGluProSer 447  
 QY 1042 CTGCTTGACCT 1053  
 Db 448 AlaLeuGluPro 451

## RESULT 12

Transcription factor HNF-3 gamma - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S35090; A39533  
 R:Lai, E.; Prezioso, V.R.; Tao, W.; Chen, W.S.; Darnell, J.J.E.  
 submitted to the EMBL Data Library, January 1993  
 A:Description: Hepatocyte nuclear factor 3a belongs to a gene family in mammals that is  
 A:Reference number: S35090  
 A:Accession: S35090  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-354 <LA1>  
 A:Cross-references: EMBL:L09648; NID:G204624; PIDN:AAA41339.1; PID:G204625  
 R:Lai, E.; Prezioso, V.R.; Tao, W.; Chen, W.S.; Darnell Jr., J.E.  
 Genes Dev. 5, 416-427, 1991  
 A:Title: Hepatocyte nuclear factor 3alpha belongs to a gene family in mammals that is ho  
 A:Reference number: A39533; MUID:191160974; PMID:1672118  
 A:Accession: A39533  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 'M', 23, 'QSDRTT', 31, 'QSDRTT', 43-44, 'T', 57-115, 'A', 117-354 <LA2>  
 A:Cross-references: GB:L09647  
 C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology  
 C:Keywords: DNA binding; nucleus; transcription regulation  
 F:119-210/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:  
 Pred. No.: 2,5e-10 Length: 354  
 Score: 348.50 Matches: 85  
 Percent Similarity: 49.16% Conservative: 32  
 Best Local Similarity: 35.71% Mismatches: 71  
 Query Match: 15.92% Indels: 50  
 DB: 1 Gaps: 6

US-10-087-080-31 (1-1209) x S35090 (1-354)

QY 252 GGTGTGGCGAGGCG 311  
 Db 82 GlyAlaGlySerGlyThrGlyGlySerAlaSerGlyTyrGlyAlaProGlyProGlyLe 101  
 QY 312 CGCGAGGCGTACGCGAGCAGCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCG 359  
 Db 101 uValHisGlyLysGluMetAlaLysGlyTyrArgArgProLeuThrHisAlaLysProPr 121  
 QY 360 CTACTCGTACATCGCGCTCATCGCCATCGCGCTCGCGCTCGCGCTCGCGCTCGCGCT 419  
 Db 121 ofySerTyrlleSerLeuThrMetAlaIleGlnAlaProGlyLysMetLeuth 141  
 QY 420 GCTGGCGGAGATCAACAGTACTCATGGCAAGTTCCCTTTTCGCGCGCGAGCTACAC 479

Db 141 rLeuSerGluIleTyGlnTrpIleMetAspLeuPheProTyTyrArgGluAsnGlnGl 161  
 QY 480 GGCTGGCGCAATCGCTCGCGCACCACTTCCTCGCTCAACGACTGCTTCGTCAGGTGCT 539  
 Db 161 nArgTrpGlnAsnSerIleArgHisSerLeuSerPheAsnAspCysPheValValAl 181  
 QY 540 GCCTGACCCCTCGCGCTCGCGCAAGCACTACTCGATCTCAACCCCAACAGCGA 599  
 Db 181 aArgSerProAspLysPro---GlyLysGlySerTyTrpAlaLeuHisProSerSerGl 200  
 QY 600 GTACACTTCGCGAGGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC----- 645  
 Db 200 yAsnMetPheGluAsnGlyCysTyrlleuArgGlnLysArgPheLysLeuGluGluL 280  
 QY 645 ----- 645  
 Db 220 sAlaLysLysGlyAsnSerAlaThrSerAlaThrArgAsnGlyThrValGlySerAlaTh 240  
 QY 646 -----AGCCACCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 671  
 Db 240 rSerAlaThrThrThrAlaAlaThrAlaValThrSerProAlaGlnProGlnProThrPr 260  
 QY 672 CGGCTCG 716  
 Db 260 oProSerGluProGluAlaGlnSerGlyGluAspValGlyGlyLeuAspCysAlaSerPr 280  
 QY 717 GCCTGCG 776  
 Db 280 oProSer---SerAlaProTyTrpPheThrGlyLeuGluLeuProGly-----GluLeu 297  
 QY 777 CGCAGCG 828  
 Db 297 sLeuAspAlaProTyTrpAsnPheAsnHisProPheSerIleAsnAsnLeuMet 314

## RESULT 13

brain factor 1 protein - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 24-Sep-1999  
 C:Accession: JH0672  
 R:Tao, W.; Lai, E.  
 Neuron 8, 957-966, 1992  
 A:Title: Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork hea  
 A:Reference number: JH0672; MUID:92265309; PMID:1350202  
 A:Accession: JH0672  
 A:Molecule type: mRNA  
 A:Residues: 1-480 <TAO>  
 A:Cross-references: GB:M87634; NID:G203134; PIDN:AAA40812.1; PID:G203135  
 A:Experimental source: brain  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 C:Keywords: DNA binding; transcription factor  
 F:162-271/Region: DNA binding #status predicted  
 F:172-263/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:  
 Pred. No.: 3.53e-10 Length: 480  
 Score: 345.00 Matches: 127  
 Percent Similarity: 40.98% Conservative: 41  
 Best Local Similarity: 30.98% Mismatches: 141  
 Query Match: 15.76% Indels: 101  
 DB: 2 Gaps: 19

US-10-087-080-31 (1-1209) x JH0672 (1-480)

QY 22 CTCTCG 81  
 Db 81 ProProGlnAlaArgGlyAlaProAlaAlaAspAspLysGly----- 95  
 QY 82 CGCGCTCCCGCGTG-----TCGCGCGCGCGGAGAC-----GACTCC 117  
 Db 96 ---ProGlnProLeuLeuLeuProProSerAlaAlaLeuAspGlyAlaLysAlaAspAla 114

118 CTGGCTCAGATGGGACTCGCGCCCAAGCCGTCGCGGGCGCGCCAGAGATACG 177  
115 LeuGlyAlaValGlyGlu-----ProGlyGlyProAlaGluLeuAlaPro 130  
178 CAGCGCGACGCGCAACAGAGTCGCGCGCGCGCGCGCGAGGAGCGATCCCGGCA 237  
131 ValGlyProAspGluLysGlyAlaGlyAlaGlyGluGlu----- 146  
238 GCAGCTGTCAGCGGTGTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 297  
147 -----LysLysGlyAlaGlyGluGlyLys 155  
298 GCG 357  
156 AspGlyGluGlyGlyGlyGly-----AspLysLysAsnGlyLysTyrGluLysPro 173  
358 CCTACTCGTACATCGCTCATCGCATCGCGCATCGCGCATCGCGCGCGCGCGCTTG 417  
174 ProPheSerTyrAsnAlaLeuIleMetMetAlaIleArgGlnSerProGluLysArgLeu 193  
418 ACGCTGGCGGAGATCAACAGGATCTCATGGCAAGTTCCCTTTTCCGCGCGAGCTAC 477  
194 ThrLeuAsnGlyIleTyrGluPheIleMetLysAsnPheProTyrTyrArgGluAsnLys 213  
478 ACGGCTCGCGCAACTCGCGTGGCGCACCACTTCGCTCAACGACTGCTTCGTCAGAGTG 537  
214 GlnGlyTyrGlnAsnSerIleArgHisAsnLeuSerLeuAsnLysCysPheValVal 233  
538 CTGCGCGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597  
234 ProArgHisTyrAspAspPro---GlyLysGlyAsnTyrTrpMetLeuAspProSerSer 252  
598 GAGTACACCTTCG 636  
253 AspAspValPheIleGlyGlyThrThrGlyLysLeuArgArgSerThrThrSerArg 272  
637 AAGCGCTCAGCCACCGC-----CGCGCGTCCCGCGCGCGCGCGCGCGCGCGCG 690  
273 AlaLysLeuAlaPheLysArgGlyAlaArgLeuThrSerThrGlyLeuThrPheMetAsp 292  
691 GCCCGCGCG-----CTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 723  
293 ArgAlaGlySerLeuTyrTrpProMetSerProPheLeuSerLeuHisPheArgAla 312  
724 -----CCG 753  
313 SerSerThrLeuSerTyrAsnGlyThrThrSerAlaTyrProSerHisProMetProTyr 332  
754 TCG 807  
333 SerSerValLeuThrGlnAsnSerLeuGlyAsnAsnHisSerPheSerThrAlaAsnGly 352  
808 TTCGCGCATCGACAGCATCTCGCGCAAG-----CCCTTCGCGCGCGCGCGCGCGCG 861  
353 LeuSerValAspArgLeuValAsnGlyGluIleProTyrAlaThrHisHisLeuThrAla 372  
862 ACG 906  
373 AlaAlaLeuAlaAlaSerValProCysGlyLeuSerValProCysSerGlyThrTyrSer 392  
907 -----CGCGCTGC-----CGCGCTGC 921  
393 LeuAsnPro-CysSerValAsnLeuLeuAlaGlyGlnThrSerTyrPhePheProHisVal 412  
922 CCGCGCTCCTCCCG 981  
412 LProHisProSer-----MetThrSerG 420  
982 GCG 1041  
420 nThrSerThrSerMetSerAlaArgAlaAlaSerSerSerThrSerProGlnAlaProSe 440  
1042 CTGCTTGACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1069

Db 440 rThrLeuProCysGluSerLeuArgPro 449

## RESULT 14

B39533

transcription factor HNF-3 beta - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 24-Jan-1992 #sequence\_revision 24-Jan-1992 #text\_change 27-Oct-1995

C/Accession: B39533

R/Author: Prezioso, V.R.; Tao, W.; Chen, W.S.; Darnell Jr., J.E.

Genes Dev. 5, 416-427, 1991

A/Title: Hepatocyte nuclear factor 3alpha belongs to a gene family in mammals that is ho

A/Reference number: A39533; PMID:91160974; PMID:1672118

A/Accession: B39533

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-458 &lt;LA1&gt;

C/Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology

F;159-249/Domain: fork head DNA-binding domain homology &lt;FHD&gt;

Alignment Scores: 3.76e-10 Length: 458

Pred. No.: 344.50 Matches: 107

Score: 344.50 Conservative: 45

Percent Similarity: 44.44% Mismatches: 147

Best Local Similarity: 31.29% Indels: 44

Query Match: 15.74% Gaps: 11

DB: 2

US-10-087-080-31 (1-1209) x B39533 (1-458)

QY 7 TTGGAGGTCTTCGTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66

Db 40 MetAsnThrTyrMetSerMetSerAlaAlaAlaMetGlySerGlySerGlyAsnMetSer 59

QY 67 GCG 126

Db 60 AlaGlySerMetAsnMetSerSerTyrValGlyAlaGlyMetSerProSerLeu 77

QY 127 GATCGGAGCTCG 186

Db 78 -----AlaGlyMetSerProGlyAlaGlyAlaMetAlaGlyMetSerGlySerAla 94

QY 187 GCGAACAACAGAGT---CGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 231

Db 95 GlyAlaAlaGlyValAlaGlyMetGlyProHisLeuSerProSerLeuSerProLeuGly 114

QY 232 ---CGCGAGCAGCTCTCGAGCGGTGTGGCGGAGCGCGCGCGCGCGCGCGCGCG 288

Db 115 GlyGlnAlaAlaGlyAlaMetGlyGlyLeuAlaProTyrAlaAsnMetAsnMetSer 134

QY 289 CCAGGCG 345

Db 135 ProMetTyrGlyGlnAlaGlyLeuSerArgAlaArgAspProLysThrTyrArgArgSer 154

QY 346 -----CGCGCCCAAGCCCTACTCGTACATCGCGCTCTCCCATGCCATCCCGGAC 399

Db 155 TyrThrHisAlaLysProTyrSerTyrIleSerLeuIleThrMetAlaIleGln 174

QY 400 TCAGCG 459

Db 175 SerProAsnLysMetLeuThrLeuSerGluIleTyrGlnThrPheMetAspLeuPhePro 194

QY 460 TTTTTCG 519

Db 195 PheTyrArgGlnAsnGlnArgTyrGlnAsnSerIleArgHisSerLeuSerPheAsn 214

QY 520 GACTGCTTCGTCAGGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579

Db 215 AspCysPheLeuLysValProArgAlaProAspLysProGlyGly-----SerPheTyr 232

QY 580 ATGCTCAACCCCAACGCGAGTACACCTTCGCGCGCGCGCGCGCGCGCGCGCG 639

Db 233 ThrLeuHisProAspSerGlyAsnMetPheGluAsnGlyCysTyrLeuArgGlnLys 252



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: March 29, 2004, 11:53:16 ; Search time 17.5 seconds  
(without alignments)  
7194.608 Million cell updates/sec

Title: US-10-087-080-31  
Perfect score: 2189  
Sequence: 1 atgaagtggagggtgttcgt.....tgagagcgtccctagattga 1209

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-Q=/cgn2.1/USPTO.spool.p/US10087080/runat.29032004.114526.6530/app.query.fasta\_1.1351  
-DB=SwissProt.42-QFMT=fastan-SUFFIX=n2p.rsp-MINMATCH=0.1-LOOPEL=0  
-LOOPEXT=0-UNITS=bits-START=1-END=1-WATRIX=blosom62-TRANS=human40.cdi  
-LIST=45-DOALIGN=200-THR SCORE=pct-THR MAX=100-THR MIN=0-ALIGN=15  
-MODE=LOCAL-OUTFMT=prc-NORM=ext-HEAPSIZES=500-MINLEN=0-MAXLEN=2000000000  
-USER=US10087080-ECGN.1.1.16@runat.29032004.114526.6530-NCPU=6-ICPU=3  
-NO MMAP-LARGQUERY-NEG SCORES=0-WAIT-DSBLOCK=100-LONGLOG  
-DEV TIMEOUT=120-WARN TIMEOUT=30-THREADS=1-XGAPOP=10-XGAPEXT=0.5-FCGAPOP=6  
-FGAPEXT=7-YGAPOP=10-YGAPEXT=0.5-DELOP=6-DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2081.5	95.1	403	FXQ1_HUMAN	Q9C009 homo sapien
2	1589	72.6	400	FXQ1_MOUSE	O70220 mus musculus
3	1458.5	66.6	392	FXQ1_RAT	O63244 rattus norv
4	558	25.5	497	FXD2_HUMAN	Q60548 homo sapien
5	543.5	24.8	440	FXGA_CHICK	Q98937 gallus gall
6	517	23.6	465	FXD1_HUMAN	Q16676 homo sapien
7	505	23.1	478	FXD3_HUMAN	Q94J05 homo sapien
8	467	21.3	465	FXD3_MOUSE	Q61060 mus musculus
9	465.5	21.3	465	FXD3_MOUSE	Q9YJ14 mus musculus
10	445.5	20.4	388	FXE3_HUMAN	Q13461 homo sapien
11	442.5	20.2	319	FXD3_CHICK	P79772 gallus gall
12	440.5	20.1	394	FXD1_HUMAN	O00358 homo sapien
13	436.5	19.9	376	FXE1_HUMAN	Q12947 homo sapien
14	431.5	19.7	553	FXE2_HUMAN	Q61572 mus musculus
15	430.5	19.7	553	FXC1_MOUSE	Q12948 homo sapien
16	429.5	19.6	553	FXC1_HUMAN	P58012 homo sapien
17	415.5	19.0	376	FXL2_HUMAN	Q8WXT5 homo sapien
18	412	18.8	416	FXD5_HUMAN	

19	407	18.6	439	1	FXD4_HUMAN	Q12950 homo sapien
20	397.5	18.2	472	1	FXH1_BRARE	Q91961 brachydanio
21	397	18.1	345	1	FXL1_HUMAN	Q12952 homo sapien
22	394.5	18.0	408	1	FXDL_HUMAN	Q9NU39 homo sapien
23	394.5	18.0	444	1	FXD4_MOUSE	Q60688 mus musculus
24	387	17.7	473	1	FXD4_HUMAN	P55317 homo sapien
25	384.5	17.6	501	1	FXC2_HUMAN	Q99958 homo sapien
26	384	17.5	494	1	FXC2_MOUSE	Q61850 mus musculus
27	371.5	17.0	457	1	FXC2_MOUSE	Q9Y261 homo sapien
28	370.5	16.9	468	1	FXC2_MOUSE	P35582 mus musculus
29	361.5	16.5	337	1	FXL1_MOUSE	Q64731 mus musculus
30	360.5	16.5	353	1	FXF1_MOUSE	Q61080 mus musculus
31	359	16.4	354	1	FXF1_HUMAN	Q12946 homo sapien
32	356	16.3	459	1	FXB3_MOUSE	P35583 mus musculus
33	356	16.3	466	1	FXD3_MOUSE	Q02361 drosophila
34	354.5	16.2	456	1	FXD3_MOUSE	F23512 rattus norv
35	348.5	15.9	354	1	FXB2_MOUSE	P32183 rattus norv
36	346.5	15.8	480	1	FXB2_MOUSE	Q64733 mus musculus
37	345	15.8	480	1	FXB2_MOUSE	Q00939 rattus norv
38	344.5	15.7	500	1	FXE2_HUMAN	Q99526 homo sapien
39	344	15.7	469	1	FXGA_HUMAN	P55318 homo sapien
40	343.5	15.7	469	1	FXGA_MOUSE	P55318 homo sapien
41	341	15.6	481	1	FXB1_HUMAN	Q60987 mus musculus
42	340	15.5	351	1	FXI1_HUMAN	Q12951 homo sapien
43	338.5	15.5	477	1	FXB1_HUMAN	P55315 homo sapien
44	335.5	15.3	353	1	FXB3_MOUSE	P35584 mus musculus
45	329.5	15.1	325	1	FXB1_MOUSE	Q64732 mus musculus

ALIGNMENTS

RESULT 1  
FXQ1\_HUMAN STANDARD; PRT; 403 AA.  
ID FXQ1\_HUMAN Q9C009; Q9NS06;  
AC Q9C009; Q9NS06;  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Forkhead box protein Q1 (Hepatocyte nuclear factor 3 forkhead homolog  
DE 1) (HNF-3/forkhead-like protein 1) (HNF-1).  
GN FOXQ1 OR HNF1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.  
RX MEDLINE=21614676; PubMed=11747606;  
RA Bieller A., Pasche B., Frank S., Glaeser B., Kunz J., Witt K.,  
Zoll B.;  
RT "Isolation and characterization of the human forkhead gene FOXQ1.";  
RL DNA Cell Biol. 20:555-561(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2127087; PubMed=11309849;  
RA Hong H.-K., Noveroske J.K., Headon D.J., Liu T., Sy M.S.,  
Justice M.J., Chakravarti A.;  
RT "The winged helix/forkhead transcription factor Foxq1 regulates  
RT differentiation of hair in satin mice.";  
RL Genesis 29:163-171(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
Diatchenko L., Marusica K., Farmer A.A., Rubin G.M., Hong L.,  
Sapich M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.P., Mullany S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield J.S.N., Krzywinski M.R., Skalska U., Marra M.A.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in the stomach,  
 CC trachea, bladder and salivary gland.  
 CC -!- SIMILARITY: Contains 1 fork-head domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; AF225950; AAK06339.1; -;  
 DR EMBL; AF153341; AAF75586.1; -;  
 DR EMBL; BC053850; AAH53850.1; -;  
 DR HSSP; O63245; 2HPH.  
 DR Genew; HGNC:20951; FOXO1.  
 DR InterPro; IPR001766; TF\_Fork\_head.  
 DR Pfam; PF00250; Fork\_head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR ProDom; PD000425; TF\_Fork\_head; 1.  
 DR SMART; SM00339; FH; 1.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS00039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation.  
 FT DNA\_BIND 119 214 FORK-HEAD.  
 FT DONAIN 13 103 ALA/GLY-RICH.  
 FT DONAIN 221 397 PRO-RICH.  
 FT CONFLICT 49 61 NSP -> QPS (IN REF. 2).  
 FT CONFLICT 60 61 PP -> TQ (IN REF. 2).  
 FT CONFLICT 386 386 S -> L (IN REF. 2).  
 FT CONFLICT 395 395 P -> S (IN REF. 2).  
 SQ SEQUENCE 403 AA; 41491 MW; EB52255AEAC6929B CRC64;

## Alignment Scores:

Pred. NO.: 1.74e-77 Length: 403  
 Score: 2081.50 Matches: 396  
 Percent Similarity: 98.51% Conservative: 1  
 Best Local Similarity: 98.26% Mismatches: 5  
 Query Match: 95.09% Indels: 1  
 DB: 1 Gaps: 1

US-10-087-080-31 (1-1209) x FOXO1\_HUMAN (1-403)

Qy 1 ATGAAGTTGAGGTGTTGTCCTCGCGCGCCAGCGGACAAAGCAGGCGAGTGACCTG 60  
 Db 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20  
 Qy 61 GAGGCGCGCGCGCAGCAGCGCGCTCCCGCTGTCCGGCGCGGAGACGACTCCCTG 120  
 Db 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40  
 Qy 121 GGCTCAGATGGGACTGCGCGGCC---AAGCGCTCCGCGCGCGCGCGCGCCAGAGATACG 177  
 Db 41 GlySerAspGlyAspCysAlaAlaAsnSerProAlaAlaGlyGlyAlaArgAspPro 60  
 Qy 178 CAGGCGCAGCGCAACAGAGTCGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 237  
 Db 61 ProGlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGluAlaAlaProAla 80

## RESULT 2

FOXO1\_MOUSE  
 ID FOXO1\_MOUSE PRT; 400 AA.  
 AC O70220;  
 DT 28-FEB-2003 (Rel. 41, Created)

Qy 238 GCAGCTCTCTCAGCGGTGGTGGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCG 297  
 Db 81 AlalaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaGlyProGlyAla 100  
 Qy 298 GCGCGCGCGGAGCGCGGAGGAGTGTGCAAGCAAGCAAGCAAGCAAGCAAGCC 357  
 Db 101 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyThrArgArgProLysPro 120  
 Qy 358 CCCTACTCGTACATCGCGCTCATCGGCATCGGCATCGGCATCGGCATCGGCATCG 417  
 Db 121 ProTySerTyrlleAlaLeulleAlaMetAlaAlaArgAspSerAlaGlyGlyArgLeu 140  
 Qy 418 ACGTGGCGGAGATCAACAGATACCTCATGGCAAGTTCCCTTTTTCGCGCGCGAGTAC 477  
 Db 141 ThrLeuAlaGluIleAsnGluTyLeuMetGlyLysPheProPhePheArgGlySerTy 160  
 Qy 478 ACGGCTGGCGCAACTCCGTGGCCCAACCTTTCGTCACAGACTCTTCCTCAAGGTG 537  
 Db 161 ThrGlyTrpArgAsnSerValArgHisLeuSerLeuAsnAspCysPheValLysVal 180  
 Qy 538 CTGGCGACCCCTCGCGCGCGCTGGGGCAAGGACAACTACTGATGCTCAACCCCAACAGC 597  
 Db 181 LeuArgAspProSerArgProTrpGlyLysAspAsnTyTrpMetLeuAsnProAsnSer 200  
 Qy 598 GAGTACACCTTCCGCCACACGGGTCTTCCCGCGCGCGCGCGCGCGCGCGCGCG 657  
 Db 201 GluTyThrPheAlaAspGlyValPheArgArgLysArgLeuSerHisArgAla 220  
 Qy 658 CCGGTCCCG 717  
 Db 221 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaPro 240  
 Qy 718 CCG 777  
 Db 241 ProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 260  
 Qy 778 GCAGCG 837  
 Db 261 AlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysPro 280  
 Qy 838 TTCGCGAGCGCTCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCG 897  
 Db 281 PheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAla 300  
 Qy 898 CCTTGGCG 957  
 Db 301 ProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeu 320  
 Qy 958 CTGCGCTCTGCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1017  
 Db 321 LeuProLeuCysAlaTyGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGlu 340  
 Qy 1018 GTGCGACG 1077  
 Db 341 ValProThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaAlaProAlaLys 360  
 Qy 1078 CCACTCCGAGGCG 1137  
 Db 361 ProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyCysProLeuArgLeuProAla 380  
 Qy 1138 GCCTTCAGGCGCGCTTAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1197  
 Db 381 AlaLeuGlnAlaAlaSerValArgArgProGlyProHisLeuProTyProValGluThr 400  
 Qy 1198 CTCCTAGCT 1206  
 Db 401 LeuLeuAla 403





QY 1147 GCGCCCTAGTCGCGCTCTGGCCGACCTGCTGACCGGTGAGACGCTCTAGCT 1206  
 Db |||||  
 373 AlalaalaalaCysGlyProGlyProHisLeuSerTyrArgValGluThrLeuLeuAla 392

RESULT 4  
 FXD2\_HUMAN STANDARD; PRT; 497 AA.  
 AC O60548;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Forkhead box protein D2 (Forkhead-related protein FKHL17) (Forkhead-  
 DE related transcription factor 9) (FHEAC-9).  
 GN FOXD2 OR FKHL17 OR FHEAC9.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98066765; PubMed=9403061;  
 RA Ernteson S., Betz R., Lagercrantz S., Larsson C., Erickson S.,  
 RA Cederberg A., Carlsson P., Enerbaeck S.;  
 RT "Cloning and characterization of fheac-9 (FKHL17), a novel kidney-  
 RT expressed human forkhead gene that maps to chromosome 1p32-p34.";  
 RL Genomics 46:78-85(1997).  
 RN [2]  
 RP REVISIONS.  
 RA Enerbaeck S.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Probable transcription factor.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Kidney specific.  
 CC -1- SIMILARITY: Contains 1 fork-head domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF042832; AAC15421.1; -.  
 DR HSPB; O63245; 2PHE.  
 DR TRANSFAC; T02485; FOXD2.  
 DR MIM; 602211; -.  
 DR GO; GO:0003700; F:transcription factor activity; TAS.  
 DR InterPro; IPR001766; TF\_Fork\_head.  
 DR Pfam; PF00250; FORK\_HEAD\_1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR ProDom; PD000425; TF\_Fork\_head; 1.  
 DR SMART; SM00339; FH\_1.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS00339; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation.  
 FT DOMAIN 90 94 POLY-ALA.  
 FT DOMAIN 101 104 POLY-ALA.  
 FT DNA\_BIND 126 217 FORK-HEAD.  
 FT DOMAIN 247 250 POLY-ALA.  
 FT DOMAIN 296 306 POLY-ALA.  
 FT DOMAIN 398 409 POLY-GLY.  
 FT DOMAIN 421 426 POLY-GLY.  
 FT DOMAIN 442 445 POLY-ALA.  
 SQ SEQUENCE 497 AA; 49007 MW; EAAFA98D216B5019 CRC64;

Alignment Scores:  
 Pred. No.: 3,328-16 Length: 497  
 Score: 558.00 Matches: 176  
 Percent Similarity: 44.75% Conservative: 37

Best Local Similarity: 36.97% Mismatches: 152  
 Query Match: 25.49% Indels: 111  
 DB: 1 Gaps: 21

US-10-087-080-31 (1-1209) x FXD2\_HUMAN (1-497)

QY 37 GGGACAAGCAGCGCAGTGCCTGAGGCGCGCGGCGCAGCGCGCGTCCCGGTG 96  
 Db |||||  
 32 GlyGlySerGlyGlyGlyLeuProAlaAsgSerGlyProArgAlaProArgAspVal 51  
 QY 97 TCGGCGCGGAGAGACGACTCCCTGGCTCAGATGGGGACTCGCGCGCAAGCCGCGG 156  
 Db |||||  
 52 LeuProHisGlyHisGlu-----ProAla 60  
 QY 157 GGGCGCGCGCCAGAGATACGAGCGCGCAGCGCGCAGCGCGAGTGCAGG 207  
 Db |||||  
 61 GluGluAlaGluAlaAspLeuAlaGluAspGluGluGluSerGlyGlyCysSerAspGly 80  
 QY 208 GGGCGCGCGCGAGGAGCGGATCCCGGCGAGCGAGTGTGCGAGCGGTGTGGCGGAGGC 267  
 Db |||||  
 81 GluProArgAlaLeuAlaSerArgGlyAlaAlaAlaAlaGlySerProGlyProGly 100  
 QY 268 GCGAGGCC-----GGGCGCGCGCGCGCGCGCGCGCGCGGAGCGGAGGT 321  
 Db |||||  
 101 AlaAlaAlaAlaArgGlyAlaAlaGlyProGlyProGlyProProSerGlyGlyAlaAla 120  
 QY 322 GCACGCGAGCAAGCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGTGCGGAGGC 381  
 Db |||||  
 121 ThrArgSer---ProLeuVal-----LysProProTyrSerTyrAlaLeu 136  
 QY 382 GCATGGCCATCCGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGAGATCAACGAGTAC 441  
 Db |||||  
 137 ThrMetAlaLeuGlnSerProLysLysArgLeuThrLeuSerGluIleCysGluPhe 156  
 QY 442 CTCATGGGCAAGTTCCTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 501  
 Db |||||  
 157 IleSerGlyArgPheProTyrTyrArgGluLysPheProAlaTrpGlnAsnSerIleArg 176  
 QY 502 CACACCTTTTCCTCAACGACTGTCTCAAGGTGTGCGCGACCCCTCGCGCGCGCGCGG 561  
 Db |||||  
 177 HisAsnLeuSerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro--- 195  
 QY 562 GCACAGGCAACTACTGTGATGCTCAACCCCAACAGCGAGTACACTTCGCGCGCGCGGTG 621  
 Db |||||  
 196 GlyLysGlyAsnTyrTrpThrLeuAspProLysAlaAspMetPheAspAsnGlySer 215  
 QY 622 TTCGG 681  
 Db |||||  
 216 PheLeuArgArgLysArgPheLys---ArgGlnProLeuProProProHisProHis 234  
 QY 682 CCGGAGGAGGCGCGCG----- 696  
 Db |||||  
 235 ProHisProHisProGluLeuLeuArgGlyAlaAlaAlaAlaGlyAspProGly 254  
 QY 697 -----GGCGTC 702  
 Db |||||  
 255 AlaPheLeuProGlyPheAlaAlaTyrGlyAlaTyrGlyTyrGlyTyrGlyLeuAlaLeu 274  
 QY 703 CCGCGC-----GCCCG 745  
 Db |||||  
 275 ProAlaTyrGlyAlaProProProGlyProAla-ProHisProHisProHisProHisPr 294  
 QY 746 -----GCATGCGTCTCG 783  
 Db |||||  
 294 ohisAlaPheAlaPheAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 314  
 QY 784 CCGCGCGGCAAGTTCCTCCAGCTCTTCGCCATCCACAGCATCTCTGC----- 829  
 Db |||||  
 314 OdlyArgAlaAlaAlaProProProGlyProProThrAlaSerValPheAlaGlyAlaGl 334  
 QY 830 -GCAAGCCCTTCG 874  
 Db |||||  
 334 ySerAlaProAlaProAlaProAlaSerGlySerGlyProGlyProGlyProAlaGlyLe 354



QY 687 GGAGGCGCGGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT-----C 740  
 Db 252 GAlaValAsp-----ProAlaAla-PheLeuProGlnProProGlnProGlnProGlnG 270  
 QY 741 GCCCGCGATGCGCT-----CGCCGCGCGCG 764  
 Db 270 InProProCysAlaTyrGlyProTyrGlyCysGlyTyrGlyLeuGlnProTyrH 290  
 QY 765 CCAGGAGGAGCGG-----CCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCAGCTC 806  
 Db 290 isProHisSerAlaLeuPheAlaPheHisPheSerProProProArgGlnProPro 310  
 QY 807 CTTGCGCATGACAGCATCTCGGCAAGCCCTTCGCGAGCCCTCGCCCTCGCGCTCGCGCTA 866  
 Db 310 laAlaProAlaGlyAla-ProAlaAlaLeuProProProProProProProProPro 329  
 QY 867 CCGCGGAGGACGCTTCAGTGGGCGCGCGCG-----CGCCGCGCGCT-----GCC 914  
 Db 330 ProArg-----ArgArgAlaProLeuLeuProAlaAlaGluLeuAla 343  
 QY 915 CGCGTTCGCGGCTCTCCCGCGCGCGCGCTGCGAGGCGCTGCGCGCTCGCGCTCGCGCTA 974  
 Db 344 ArgThrProPheGlyTyrProHisProLeu-----GlyProAlaLeuAlaSerLeu 361  
 QY 975 CGCGCGGCGCGCGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1034  
 Db 362 HisAlaAlaLeuProGlySer-----GlyAlaAlaValAlaArg 374  
 QY 1035 GCCCTCTCTGTCACCTCT--CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1082  
 Db 375 SerProPheSerIleGluSerIleIleGlyGlyProGlyProGlyLeuGlyAlaGly 394  
 QY 1083 CCGAGCT 1133  
 Db 395 ProAlaProGlyAlaGlySerCysAlaSerGlnSerGlyAlaAlaThrGlyLeuSer 414  
 QY 1134 CGCAGCGCTGCGAGCGCGCTGTCGCGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCT 1190  
 Db 415 ArgSerLeuGlySerGlyLeuAlaProAlaAla--AlaLeuProAlaAlaProGly 432

RESULT 6  
 FXD1\_HUMAN STANDARD; PRT; 465 AA.  
 AC Q16676; Q12949;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Forkhead box protein 1 (Forkhead-related protein FKHL8) (Forkhead-  
 related transcription factor 4) (FREAC-4).  
 GN FOXD1 OR FKHL8 OR FREAC4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96355467; PubMed=8702877;  
 RA Ennsson S., Bierrou S., Hulander M., Cederberg A., Hellqvist M.,  
 RA Carlsson P., Enerback S.;  
 RT "Characterization of the human forkhead gene FREAC-4. Evidence for  
 RT regulation by Wilms' tumor suppressor gene (WT-1) and p53."  
 RL J. Biol. Chem. 271:21094-21099(1996).  
 RN [2]  
 RP SEQUENCE OF 120-225 FROM N.A.  
 RX MEDLINE=95045392; PubMed=7957066;  
 RA Bierrou S., Hellqvist M., Samuelsson L., Enerback S., Carlsson P.;  
 RT "Cloning and characterization of seven human forkhead proteins:  
 RT binding site specificity and DNA binding."  
 RL EMBO J. 13:5002-5012(1994).  
 CC -/- FUNCTION: Binding of FREAC-3 and FREAC-4 to their cognate sites  
 CC results in location of the DNA at an angle of 80-90 degrees.  
 CC -/- SUBCELLULAR LOCATION: Nuclear.  
 CC -/- SIMILARITY: Contains 1 fork-head domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC ENBL; US9832; AAC50661.1; -  
 CC ENBL; US9831; AAC50660.1; -  
 CC ENBL; UI3222; AAA92039.1; -  
 CC PIR; G02738; G02738.  
 CC PIR; S51627; S51627.  
 CC HSP; O63245; 2HPH.  
 CC TRANSFAC; T02472; -  
 CC Genew; HGNC:3802; FOXD1.  
 CC MIM; 601091; -  
 CC GO; GO:0003700; F:transcription factor activity; TAS.  
 CC InterPro; IPR001766; TF\_Fork\_head.  
 CC Pfam; PF00250; Fork\_head; 1.  
 CC PRINTS; PR00053; FORKHEAD.  
 CC ProDom; PD000425; TF\_Fork\_head; 1.  
 CC SMART; SM00339; FH; 1.  
 CC PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 CC PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 CC PROSITE; PS00659; FORK\_HEAD\_3; 1.  
 CC Transcription regulation; DNA-binding; Nuclear protein.  
 KW DOMAIN 26 34 POLY-GLU.  
 FT DOMAIN 39 43 POLY-GLY.  
 FT DOMAIN 52 57 POLY-ARG.  
 FT DOMAIN 69 72 POLY-GLU.  
 FT DOMAIN 73 76 POLY-ASP.  
 FT DOMAIN 97 113 POLY-GLY.  
 FT DOMAIN 124 215 FORK-HEAD.  
 FT DOMAIN 231 234 POLY-ALA.  
 FT DOMAIN 252 256 POLY-ALA.  
 FT DOMAIN 259 266 POLY-PRO.  
 FT DOMAIN 293 303 POLY-ALA.  
 FT DOMAIN 309 315 POLY-PRO.  
 FT DOMAIN 375 390 POLY-ALA.  
 FT DOMAIN 428 434 POLY-ALA.  
 SQ SEQUENCE 465 AA; 46140 MW; D3E7854909CCBF AE CRC64;  
 Alignment Scores:  
 Pred. No.: 1.5e-14 Length: 465  
 Score: 517.00 Matches: 157  
 Percent Similarity: 44.21% Conservative: 34  
 Best Local Similarity: 36.34% Mismatches: 132  
 Query Match: 23.62% Indels: 109  
 DB: 1 Gaps: 16  
 US-10-087-080-31 (1-1209) x FXD1\_HUMAN (1-465)  
 QY 60 GGAGGCTCCCT 119  
 Db 39 GlyGlyGlyGlyGlyGlyProArgLeuAlaValPro----- 49  
 QY 120 GGGCTCAGATGGGGACTCG 179  
 Db 50 -----AlaGlnArgArgArgArgArgArgArgArgArgArgArgArgArgArg 60  
 QY 180 GGGCGACG 203  
 Db 61 GlyGluAspGluLeuGluAspLeuGluGluGluGluGluGluGluGluGluGluGlu 80  
 QY 204 AGCG 263  
 Db 81 ProProAlaGlyGlySerProAlaProGlyProAlaProAlaGlyGlyGlyGlyGly 100  
 QY 264 GGGCG 323  
 Db 101 GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 119

```

QY 324 ACCGACAGCCATATACGGCGGCGCCCAAGCCCTCTCTCTATCATCGCCCTCATCCG 383
Db 119 aLYsAsn---ProLeuVal-----LysProProTyrSerTyrIleAlaLeuIlePh 135
QY 384 CATGGCCATCGCGACTCGCGCGCGCGCTTACGCTGGCGGAGATCAACGAGTACCT 443
Db 135 rMetAlaIleLeuGlnSerProLYsLYsArgLeuThrLeuSerGluIleCysGluPheI 155
QY 444 CATGGCAGATTCCTCCCTTTTCGCGCGGAGCTACACGGGCTGCGGCACTCGTGGCCA 503
Db 155 eSerGlyArgPheProTyrTyrArgGluLYsPheProAlaTyrGlnAsnSerIleArgH 175
QY 504 CAACCTTCCTCAACGAGCTCTCTCAAGGTCTGCGCGAGCTCCGCGCCCTCGGG 563
Db 175 sAnLeuSerLeuAsnAspCysPheValLYsIleProArgGluProGlyAsnPro---Gl 194
QY 564 CAAGGACAATCTAGTATGCTCAACCCCAACAGCGAGTACACTTCGCGCAGCGGTCT 623
Db 194 yLYsGlyAsnTyrTrpThrLeuAspProGluSerAlaAspMetPheAsnGlySerPh 214
QY 624 CCGCCGCGCGCAAGCGCTCAGCCAGCGCGCGCGCTCCCGCGCGCGCTCGGCC 683
Db 214 eLeuArgArgLYsArgPheLYs---ArgGlnProLeuLeuProProAsnAlaAla 233
QY 684 CGAGGAGGCC-----CCGCGCTCCCGCGCGC 710
Db 233 aAlaGluSerLeuLeuArgGlyAlaGlyAlaAlaGlyGlyAlaGlyAspProAlaAl 253
QY 711 CCGG-----CCGCGCGCGCGCGCGCGCGCGCTCGC----- 742
Db 253 aAlaAlaLeuPheProProAla-ProProProProProHisAlaTyrGlyTyrGlyP 273
QY 743 -----CCGCGCTCGCTCCCGCGCGCGCGCGCGAGGA 773
Db 273 roTyrGlyCysGlyTyrGlyLeuGlnLeuProTyrAlaProProSerAlaLeuPheA 293
QY 774 CGCGCGCAGCGCGCGGCGAGTCTCCAGCTCTCGCATCGCAGCATCTCGCGCA 833
Db 293 laAlaAlaAlaAlaAlaAlaAlaAlaAlaPheHisProHis-----S 308
QY 834 GCGCTTCGCGCGCGCTCGCTCAGGAGCAGCGCGCGCGCGCGCTTCAGTGGCGCG 893
Db 308 erPro-ProProProProProPro---HisGly----- 317
QY 894 CGCGCGCTCGCGCGCTCGCGCGCTTCCTCCCGCGCT-----CCT 932
Db 318 -----AlaAlaAlaGluLeuAlaAlaArgThrAlaPheGlyTyrArgProHisPro 333
QY 933 CCGCGCGCGCGCTCGCGCGCGCTCGCTCGCGCTCGCGCGCGCGCGCGCGCGCG 992
Db 334 LeuGlyAlaAlaLeuProGlyProGlyProLeuProAlaSerAlaAlaLYsAlaGlyGlyProGly 353
QY 993 CGCGCTCGCGCGCGCGCGCGCG-----CGAGTGCACCGCAC 1028
Db 354 AlaSerAlaLeuAlaArgSerProPheSerIleGluSerIleIleGlySerLeuGly 373
QY 1029 CGCGCGCGCGCTCTCTGTGTGACCTCTCCGCGCGCGCGCGCGCGCGCGCGCG 1088
Db 374 ProAlaAlaAlaAlaAlaAlaAlaAlaAlaGlnAlaAlaAlaAlaAlaSerProSer 393
QY 1089 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1148
Db 394 Pro-----SerProValAlaAlaProAlaProAlaProGlySerSerGlyGly 408
QY 1149 GGCCTTAGTCGNGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1178
Db 409 GlyCysAlaAlaGlnAlaAlaAlaValGlyPro 418

```

RESULT 7

FXD1\_MOUSE

ID FXD1\_MOUSE

AC Q61345;

```

DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Forkhead box protein D1 (Forkhead-related protein FKHL8) (Forkhead-
DE related transcription factor 4) (FREAC-4).
GN FOXD1 OR FKHL8 OR FREAC4 OR FHBF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Embryo;
RX MEDLINE=95114592; PubMed=7815060;
RA Hatini V., Tao W., Lai E.;
RT "Expression of winged helix genes, BF-1 and BF-2, define adjacent
RT domains within the developing forebrain and retina.";
RL J. Neurobiol. 25:1293-1309(1994).
CC -!- FUNCTION: May function in specifying positional identity in the
CC developing retina as well as the subdivision of the forebrain
CC neuroepithelium.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: Predominantly expressed in the CNS and
CC temporal half of the retina. Also expressed in the condensed head
CC mesenchyme, metanephric blastema of the developing kidney, cortex
CC of the adrenal gland, condensed mesenchyme at the base of the
CC follicles of vibrissae and cartilage perichondrium of the
CC developing vertebrate.
CC -!- DEVELOPMENTAL STAGE: At E9.5 embryos, expressed in a limited
CC region of the neuroepithelium and also in the temporal half of the
CC primary optic cup and the optic stalk. At E10.5, seen in the
CC hypothalamus, temporal half of the optic stalk, and temporal
CC hemirhina. At E12.5 and E13.5 a high expression is seen in
CC regions of condensed mesenchyme of the head, and as
CC neuroepithelial cells begin to differentiate and migrate outward
CC from the ventricular zone, expression declines markedly. By E16.5
CC levels are diminished and restricted to unused pockets along the
CC exhausted ventricular zone.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC -!- CAUTION: Was originally (Ref.1) assigned to be BF-2 (FOXG1A).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; L38607; AAC42042.1; -.
CC TRANSFAC; T02293; -.
CC MGD; MGI:1347463; Foxd1.
CC InterPro; IPR001766; TF_Fork_head.
CC Pfam; PF00250; Fork_head; 1.
CC PRINTS; PR00053; FORKHEAD.
CC ProDom; PD000425; TF_Fork_head; 1.
CC SMART; SM00339; FH; 1.
CC PROSITE; PS00657; FORK_HEAD_1; 1.
CC PROSITE; PS00658; FORK_HEAD_2; 1.
CC PROSITE; PS50039; FORK_HEAD_3; 1.
CC Transcription regulation; DNA-binding; Nuclear protein;
CC Developmental protein.
CC DOMAIN 29 32 POLY-GLU.
CC DOMAIN 33 36 POLY-ASP.
CC DOMAIN 38 46 POLY-GLY.
CC DOMAIN 55 58 POLY-ARG.
CC DOMAIN 73 76 POLY-ASP.
CC DOMAIN 130 224 FORK-HEAD.
CC DNA_BIND 261 267
CC DOMAIN 293 302 POLY-PRO.
CC DOMAIN 308 319 POLY-ALA.
CC DOMAIN 395 400 POLY-GLY.

```



```

CC -----
DR ENBL; AF197560; AAF05844.1; -.
DR ENBL; AL049636; CAC19431.1; -.
DR ENBL; LI2142; AAK13574.1; -.
DR HSP; O63245; 2HPH.
DR TRANSFAC; T04167; -.
DR Genew; HGNC; 3804; FOXD3.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0016563; F:transcriptional activator activity; NAS.
DR GO; GO:0007275; P:development; NAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK HEAD 1; 1.
DR PROSITE; PS00658; FORK HEAD 2; 1.
DR PROSITE; PS00659; FORK HEAD 3; 1.
KW DNA-binding; Nuclear protein; transcription regulation.
FT DNA_BIND 141 235 FORK-HEAD.
FT DOMAIN 262 288 POLY-ALA.
FT DOMAIN 280 285 POLY-ALA.
FT DOMAIN 290 300 POLY-ALA.
FT DOMAIN 350 355 POLY-ALA.
FT DOMAIN 460 473 POLY-ALA.
FT CONFLICT 200 200 V -> D (IN REF. 3).
FT CONFLICT 207 210 GNPG -> ATRP (IN REF. 3).
FT CONFLICT 239 239 R -> A (IN REF. 3).
SQ SEQUENCE 478 AA; 47630 MW; 06163F6318FB13CB CRC64;

Alignment Scores:
Pred. No.: 1.53e-12 Length: 478
Score: 467.00 Matches: 161
Percent Similarity: 39.88% Conservative: 36
Best Local Similarity: 32.59% Mismatches: 164
Query Match: 21.33% Indels: 134
DB: 1 Gaps: 17

US-10-087-080-31 (1-1209) x FXD3_HUMAN (1-478)
QY 37 GGGGACAGCAGGCGGCTGACCTGGAGGC-----GCGGGCGGACGACGCGCG 87
Db 5 GlyGlySerAlaSerMetSerGlyGlnThrValLeuThrAlaLysValAsp 24
QY 88 TCCCGCGTGTGCGCGCGGAGACGACTCCCTGGGTCTAGATGGGACTGCGCGCCCAAG 147
Db 25 IleAspValValGlyGlyGlyAspGlyLeuGluGluLysAspSerAspAlaGlyCys 44
QY 148 CCTCCGCGCGCGCGCGCCAGATACGAGGCGGCGGCGGACAGAGTG----- 199
Db 45 AspSerProAlaGly-ProProGluLeuArgLeuAspGluAlaAspGluValProAla 64
QY 200 -----CGGA 204
Db 64 aAlaProHisHisGlyGlnProGlnProProHisGlnGlnProLeuThrLeuProLysG 84
QY 205 GCGGCGCGCGCGCGA-----GGAGCGCAT-----CCGCGCAGCAGTGTCT 246
Db 84 uAla-AlaGlyAlaGlyAlaGlyProGlyGlyAspValGlyAlaProGluAlaAspGlyC 104
QY 247 GCAGCGTGTGTGCGGAGGCGCG-----GGAGCGCGCGCGCGGCGGCGGCGCGG 300
Db 104 yslsGlyGlyValGlyGlyGluGluGlyGlyAlaSerGlyGly-GlyProGlyAlaGly 123
QY 301 GCGCGCGGAGCGCGGAGGTGACGACAGCAAGCCATATACGCGGCGCGCGCGCGCGCC 360
Db 124 SerGlySerAlaGlyGlyLeuAlaProSerLysProLysAsnSerLeuValLysPro 143
QY 361 TACTCTACATCGGCTCATCTGCGCATCGGACTCGGCGGCGGCGGCTTGAG 420
Db 144 TyrSerTyrIleAlaLeuThrMetAlaIleLeuGlnSerProGlnLysLysLeuThr 163

```

## RESULT 9

FXD3 MOUSE

ID FXD3 MOUSE

AC Q61060;

STANDARD; PRT; 465 AA.

```

QY 421 CTGGCGGAGATCAACGAGTACCTCATGGGCAAGTCCCTTTTCCGGCGGACCTACACG 480
Db 164 LeuSerGlyIleCysGluPheIleSerAsnArgPheProTyrTyrArgGluLysPhePro 193
QY 481 GCGTGGCGCAACTCCGTGCGCCACCAACCTTTCCTCAACGAGCTGCTTCGTCAGGTCGTG 540
Db 184 AlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPheValIlePro 203
QY 541 CGGACCCCTCGGGCGCTGGGCGAGGACAACTACTGATGTCTCAACCCACACGAGGAG 600
Db 204 ArgGluProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLeuAspProGlnSerGlu 222
QY 601 TACACCTTCGCGACGCGGTCTTCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGGCC- 659
Db 223 AspMetPheAspAsnGlySerPheLeuArgArgArgLysArgPheLysArgHisGlnGln 242
QY 660 -----GGTCCCCGCGCGCGGCT 677
Db 243 GluHisLeuArgGluGlnThrAlaLeuMetMetGlnSerPheGlyAlaTyrSerLeuAla 262
QY 678 GCGCGCGGAGGCGCGCGGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737
Db 263 AlaAlaAlaGlyAlaAlaGlyProTyrGlyArgPro-----TyrGlyLeuHisProAla 280
QY 738 CTCGCCCGCATGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 797
Db 281 AlaAlaAlaGlyAlaTyrSerHisProAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300
QY 798 CTCGAGCTCCTTCGCCATCGACAGCAGCTCTCGCAAGCCCTTCGCGAGCGGTG----- 851
Db 301 LeuGlnTyrProTyrAlaLeuProProValAlaProValLeuPro---ProAlaValPro 319
QY 852 -----CCTCAGGACACGCGCGCGCGCGCGCTTACGTGGGCGCGCGC----- 896
Db 320 LeuLeuProSerGlyGluLeuGlyArgLysAlaAlaAlaPheGlySerGlnLeuGlyPro 339
QY 897 -----GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 920
Db 340 GlyLeuGlnLeuGlnLeuAsnSerLeuGlyAlaAlaAlaAlaAlaAlaGlyThrAlaGly 359
QY 921 -----CCCCGCGCTCTCCCCGCGCGCGC 944
Db 360 AlaAlaGlyThrAlaSerLeuLeuLysSerGluProSerAlaArgProSerPheSer 379
QY 945 CTGCAG-----GGCCCTGTCGCGCTCTGCGCTAGCGGCGCGCGCGCGCGCGCG 989
Db 380 IleGluAsnIleIleGlyGlyGlyProAlaAlaPro-----GlyGlySerAla 395
QY 990 GCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1049
Db 396 ValGlyAlaGlyValAlaGlyGlyThrGlyGly----- 406
QY 1050 ACCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1106
Db 407 ---SerGlyGlyGlySerThrAlaGlnSerPheLeuArgProGlyThrValGlnSer 425
QY 1107 -----GCA 1109
Db 426 AlaAlaLeuMetAlaThrHisGlnProLeuSerLeuSerArgThrThrAlaThrIleAla 445
QY 1110 CCTGTACTG-----CCCTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1139
Db 446 ProIleLeuSerValProLeuSerGlyGlnPheLeuGlnProAlaAlaSerAlaAlaAla 465
QY 1140 CCTCAGCGCGCGCTTGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1175
Db 466 AlaAlaAlaAlaAlaGlnAlaLysTrpProAla 477

```

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Forkhead box protein D3 (HNF3/FH transcription factor genesis)  
 DE (Hepatocyte nuclear factor 3 forkhead homolog 2) (HFN-2).  
 GN FOXD3 OR FHF2  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Hromas R.A., Costa R.H., Xu D., Sutton J.L.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RA Labosky P.A., Kaestner K.H.;  
 RT "The winged helix transcription factor Fhf2 is expressed in neural  
 RT crest and spinal cord during mouse development."  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Probable transcription factor.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: Contains 1 fork-head domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U41047; AAA87569.1; -;  
 DR EMBL; AF067421; AAC28352.1; -;  
 DR HSSP; O63245; 2HFH.  
 DR TRANSFAC; T04166;  
 DR MGD; MGI:1347473; Foxd3.  
 DR InterPro; IPR001766; TF\_Fork\_head.  
 DR Pfam; PF00250; Fork\_head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR ProDom; PD000425; TF\_Fork\_head; 1.  
 DR SMART; SM00339; FH; 1.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation.  
 FT DOMAIN 106 115  
 FT POLY-GLY.  
 FT DNA\_BIND 131 225  
 FT FORK-HEAD.  
 FT DOMAIN 252 257  
 FT POLY-ALA.  
 FT DOMAIN 285 270  
 FT POLY-ALA.  
 FT DOMAIN 275 281  
 FT POLY-ALA.  
 FT DOMAIN 380 399  
 FT POLY-GLY.  
 FT DOMAIN 447 457  
 FT POLY-ALA.  
 SQ SEQUENCE 465 AA; 47092 MW; 6F8B5BB3D8C7564D CRC64;

## Alignment Scores:

Pred. No.: 1,778-12 Length: 465  
 Score: 465.50 Matches: 154  
 Percent Similarity: 39.17% Conservative: 34  
 Best Local Similarity: 32.08% Mismatches: 173  
 Query Match: 21.27% Indels: 119  
 DB: 1 Gaps: 16

US-10-087-080-31 (1-1209) x FOXD3\_MOUSE (1-465)

QY 37 GGGGACAAGCAGGGCAGTGAAGGCG-----GGCGGGGAGCAGCAGCGCGCG 87  
 Db |||||:|  
 D 5 GlySerGlySerAlaSerAspMetSerGlyGlnThrValLeuThrAlaGluAspValAsp 24  
 QY 88 TCCCGCTCTCGCGGGGAGACGACCTCCG-----GGCTCAGATGGGACTGC 138  
 Db |||||:|  
 D 25 IleAspValValGlyGluGlyAspGlyLeuGluGluLysAspSerAspAlaGlyCys 44

QY 139 -----CGCGCCAAAGCGG----- 150  
 D 45 AspSerProAlaGlyProProAspLeuArgLeuAspGluAlaAspGluGlyProProVal 64  
 QY 151 TC CGCGCGGGCGGGCGCCAGAGATACGAGGGCGACGCGGCGAAACAGAGTGGGAGCGGG 210  
 D 65 SerAlaHisHisGlyGlnSerGlnProGlnAlaLeuAlaLeuProThrGluAlaThrGly 84  
 QY 211 CGCGCGCGGAGGAGCGATCCCGGAGCAGCTGTGTCAGCGGTGCTGGCGGAGGCGGG 270  
 D 85 ProGlyAsnAspThrGlyAlaProGluAlaAspGlyCysLys---GlyGlyGluAspAla 103  
 QY 271 GAGCGCGGGCGCGGGCGGCGAGCGCGGGCGGAGGAGGCGGAGGTGCACGACG 330  
 D 104 ValThrGlyGlyGlyProGlyAlaGlySerGlyValaThrGlyGlyLeuThrProAsn 123  
 QY 331 AAGCATATACGGCGGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 390  
 D 124 LysProGlyAsnSerLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAla 143  
 QY 391 ATCCGCGACTCGCGGGCGGGCGCTTACGCTGCGCGGAGTCAACGAGTACCTCATGGGC 450  
 D 144 IleLeuGlnSerProGlnLysLysLeuThrLeuSerGlyIleCysGluPheIleSerAsn 163  
 QY 451 AAGTTCCTCTTTTCGCGGCGAGTACAGGGGTGGCGCACTCCGTGGCGCCACACCTT 510  
 D 164 ArgPheProTyrTyrArgGluLysPheProAlaIleArgGlnAsnSerIleArgHisAsnLeu 183  
 QY 511 TCCTCAAGACGCTTCGTCAGGNGTCGCGGAGCCCTCGCGGCGCTGGGCGCAAGGAC 570  
 D 184 SerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro---GlyLysGly 202  
 QY 571 AACTACTGGATGCTAACCCCAACAGCGAGTACACTTCGCGGACGGGTCTTCGCGCGC 630  
 D 203 AsnTyrTrpThrLeuAspProGlnSerGluAspMetPheAspAsnGlySerPheLeuArg 222  
 QY 631 CGCGCGCAAGCGC----- 642  
 D 223 ArgArgLysArgPheLysArgHisGlnGlnGluHisLeuArgGluGlnThrAlaLeuMet 242  
 QY 643 -----CTCAGCCACCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 678  
 D 243 MetGlnSerPheGlyAlaTyrSerLeuAlaAlaAlaGlyAlaGlyProTyrGlyLeu 262  
 QY 679 CGCGCGGAGAGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 738  
 D 263 HisProAlaAlaAlaAlaGly-----AlaTyrSerHisProAlaAlaAlaAlaAla 280  
 QY 739 TCGCCCGCGCATGCGCTCGCGCG-----CCCGC 765  
 D 281 Ala-GlyCysCysGlyAlaProValProValArgAlaThrThrArgGlySerArgAla 300  
 QY 766 CAGGAGAGCGCGCGCGCG-----CCGCGGGCAAGTCTCCAGCTCC 807  
 D 300 aAlaArgSerAlaProAlaAlaLeuGlyArgAlaGlyProGlnSerGlyArg-LeuArg 320  
 QY 808 TTCGCCCATCGACAGCATCTCGGCAAGCCCTTCGCGAGCGG-----TCGCCTC 855  
 D 320 euAlaAlaArgSerGluProAlaAlaAlaThrAlaGlnHisProGlyArgArgArgGlySerA 340  
 QY 856 AGGACACGCGCGCGCGCGCGCGCTTCAGTGGGCGC-----CGCGCGCTCGCGCGCG 909  
 D 340 rgGlyHisGlyGlyArgGlyGlyHisHisValAlaHisGlnValArgAlaGlnCysAlaA 360  
 QY 910 CTCGCCCGCTTCCCGCGCTCTCCCGCGCGCGCTGCAGGGCGCTGTCGCGCTCTGC 969  
 D 360 laValValGlnHisArgGluHisHisArgGlyIleArgAlaProGlyGlySer----- 378  
 QY 970 GCGTACGCGCGCGCGCGCGCGCGCTGGCGCGCGCGCGCGCGCGAGCGAGTGCACGACC 1029  
 D 379 -----AlaGlyGlyGlyGlySerGlyGlyAlaGlyGlyGlyGlyGly----- 393  
 QY 1030 GCGCGCGCGCTCTGCTTGACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1089



```

QY 1003 GCGCGGAGCGGAGTGCACG-----ACGCGCGCCCTCTCTGCTTGCACCT 1053
Db 220 AlaProAspAlaAlaPheProProCysAlaAlaAlaSerProProLeuTyrSerPro 239
QY 1054 -----CTCCGCGGCGGCGCGCCCGCGCAAGCCACTCCGAGGCGCGCGG 1095
Db 240 AlaserGluArgLeuGlyLeuProAlaProLeuProAlaGlnProLeuAlaA---Leu 258
QY 1096 GCGCGGCGGCGGACCTGTACTGCTCCCTCTGCGGCTCGCGAGCCCTGAGCGGCGCTTA 1155
Db 259 AlaGlySerAlaGlyAlaLeuGlyProLeu-----GlyAlaGlyGluAlaTyr 274

RESULT 11
FXE3_HUMAN STANDARD; PRT; 319 AA.
ID FXE3_HUMAN AC Q13451; Q9NOV9;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Forkhead box protein E3 (forkhead-related protein FKHL12) (Forkhead-
GN FOXE3 OR FKHL12 OR FREAC8).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Carlsson P.;
RT "Forkhead transcription factor FOXE3 is expressed in lens
RT epithelium";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 66-171 FROM N.A.
RX MEDLINE:96423031; PubMed:8825632;
RA Larsson C., Hellqvist M., Pierrou S., White I., Enerbaeck S.,
RA Carlsson P.;
RT "Chromosomal localization of six human forkhead genes, freac-1
RT (FKHL5), -3 (FKHL7), -4 (FKHL8), -5 (FKHL9), -6 (FKHL10), and -8
RT (FKHL12).";
RL Genomics 30:464-469(1995).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275722; AAF82793.1; --
DR EMBL; U42990; AAB48856.1; --
DR PIR; G02311; G02311.
DR HSSP; Q63245; 2HPH.
DR TRANSFAC; T02476; --
DR Genew; HGNC:3808; FOXE3.
DR MIM; 601094; --
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007275; P:development; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.

```

```

DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00339; FORK_HEAD_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 61 65 POLY-ARG.
FT DNA BIND 70 161 FORK-HEAD.
FT DOMAIN 174 179 POLY-ALA.
FT DOMAIN 245 249 POLY-ALA.
FT DOMAIN 254 257 POLY-ALA.
FT CONFLICT 162 162 R -> P (IN REF. 2).
SQ SEQUENCE 319 AA; 33234 MW; E25A64457B7ECDP8 CRC64;

Alignment Scores:
Pred. No.: 1.58e-11 Length: 319
Score: 442.50 Matches: 134
Percent Similarity: 43.85% Conservative: 30
Best Local Similarity: 35.83% Mismatches: 129
Query Match: 20.21% Indels: 81
DB: 1 Gaps: 14

US-10-087-080-31 (1-1209) x FXE3_HUMAN (1-319)
QY 67 GCGGCGGCGGAGCGCGCGCTCCCGCTGTCGCGCGGCGGAGCAGTCCCTGGGCTCA 126
Db 2 AlaGlyArgSerAspMetAspProProAlaAlaPheSerGlyPheProAlaLeuProAla 21
QY 127 GATGGGACTGCGCGGCCAAGCCGTCCGCGGCGGCGGCCAGAGATACGACGGCGAC 186
Db 22 ValAlaProSerGlyProProSerProLeuAlaGlyAlaGluProGlyArgGluPro 41
QY 187 GCGAAGACAGTGGGAGGCGGCGGCGGCGGAGGAGGCGATCCCGGCGAGCTGCT 246
Db 42 GluGluAlaAlaGlyArgGly----- 49
QY 247 GCAGCGTGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 306
Db 50 -----GluAlaAlaProThrProAlaProGlyProGlyArg----- 61
QY 307 GGGAGCGCGAGGGTGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 366
Db 62 -----ArgArgArgArgPro--LeuGlnArgGlyLysProProTyrSer 75
QY 367 TACATCGCTCATGCGCATGCGCATGCGCATGCGCATGCGCATGCGCATGCGCATGCG 426
Db 76 TyrIleAlaLeuIleAlaMetAlaLeuAlaHisAlaProGlyArgLeuThrLeuAla 95
QY 427 GAGATCAACGAGTACCTCATGCGCAAGTTCCTCCCTTTTCGCGGCGAGCTACAGGCTGG 486
Db 96 AlaIleTyrArgPheIleThrGluArgPheAlaPheTyrArgAspSerProArgLysTrp 115
QY 487 CGCAACTCCGTGCGGCACACACTTCGCTCAAGACTGCTTCGCTCAAGGTGCTGCGGAC 546
Db 116 GlnAsnSerIleArgHisLeuThrLeuAsnAspCysPheValLysValProArgGlu 135
QY 547 CCCTCGCGGCGCTGGGCGAAGCAACTACTGGATGCTCAACCCCAACAGCGAGTACACC 606
Db 136 ProGlyAsnPro--GlyLysGlyAsnTyrTrpThrLeuAspProAlaAlaAspMet 154
QY 607 TTGCGGACGCGGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 648
Db 155 PheAspAsnGlySerPheLeuArgArgArgLysArgLysArgAlaGluLeuProAla 174
QY 649 CACGCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708
Db 175 HisAlaAlaAlaProGlyPro-----ProLeuProPheProTyrAlaProTyr 191
QY 709 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 768
Db 192 AlaProAlaProGlyProAlaLeuLeuValProProSerAlaGlyProGly----- 209
QY 769 GAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828
Db 210 -----ProSerProProAlaArgLeu-----PheSerValAspSerLeuVal 223

```

```
QY 829 CGAAGCCCTTCGCGACCGCTCCAGGACACAGCCCGCCCGGAGCAGCGTTCAGTGG 888
Db 224 -----AsnLeuGlnProGluLeuAlaGlyLeu----- 232
QY 889 GGGCGCGCGCCCTGCGCGCGCTGCGCGCTTCCCGCGCTCTCCCGCGCGCCCTGC 948
Db 233 ---GlyAlaProGluProCysCysAlaProAspAlaAlaAlaAlaAlaAlaAla 249
QY 949 AGGCGCCCTGCTGCGCTCTCGCGGTACGGCGCGCGGAGCGCGCGCTGGCGCGCGC 1008
Db 250 -----PheProCysAlaAlaAlaAlaAlaSerProLeu----- 261
QY 1009 GAGCGCAGGTGACCGACCGCGCGCGCCCTCTCTGCTTGCACCTCTCCCGCGCGCGC 1068
Db 262 TyrSerGlnValPro---AspArgLeuValLeuProAla---ThrArgProGlyProGlyP 280
QY 1069 CCGGCCAAGCACTCCGAGGCGCGCGCGCGCGCGCGCGCTGACTGCGCCCTGCGG 1128
Db 280 ro-----LeuProAlaGluProLeuLeuA 288
QY 1129 CTGCGCGCAGCCCTGACGGCGGCTTAGTCCGNGCTCC 1166
Db 288 laLeuAlaGlyProAlaAlaAlaLeuGlyProLeuSer 300

RESULT 12
FXD3_CHICK STANDARD; PRT; 394 AA.
AC P97772;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein D3 (HNF3/FH transcription factor gene) (Winged
DE helix protein CW3-3).
GN FOXD3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=97141794; PubMed=8988052;
RA Freyaldenhoven B.S., Freyaldenhoven M.P., Iacovoni J.S., Vogt P.K.;
RT "Aberrant cell growth induced by avian winged helix proteins.";
RL Cancer Res. 57:123-129(1997).
CC -!- FUNCTION: Probable transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis/announcement/
CC or send an email to license@ebi.ac.uk).
```

```
FT DNA BIND 117 211 FORK-HEAD.
SQ SEQUENCE 394 AA; 40995 MW; 324A4B36B9E31899 CRC64;

Alignment Scores:
Pred. No.: 1,84e-11 Length: 394
Score: 440.50 Matches: 147
Percent Similarity: 41.52% Conservative: 39
Best local Similarity: 32.81% Mismatches: 122
Query Match: 20.12% Indels: 140
DB: 1 Gaps: 19

US-10-087-080-31 (1-1209) x FXD3_CHICK (1-394)
QY 49 GCGAGTACCTGGAGGCG-----GCGGCGGACGAGCGCGCTCCCGCTCGTCG 99
Db 7 GlySerAspMetSerGlyGlnThrAlaLeuAlaGluAspValAspIleAspVal 26
QY 100 GCGGCGGAGACGACTCCCTCGGCTCAGATGGGACTGCGCGGCGCAAGCCGCTCCGCGGCG 159
Db 27 GlyGluGlyAspAlaProGlyLysAspGlyGluAlaArg-SerProAlaAla 46
QY 160 G-----GCGGCGGACGAGATACGAGGGCGAC 186
Db 46 aLeuProLeuProLeuAspGluAlaAlaGluProGlyGluProGlu---ArgAlaAlaA 65
QY 187 GCGCAACAGAGTGGCGGAGCGCGCGCGCGGAGGAGGCGATCCCGGACGAGCTGCT 246
Db 65 rGArgAlaAlaAlaAlaArgGlnProGlyArgProGluGlyGlyArg----- 82
QY 247 GCAGCGTGTGGCGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 83 --GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 101
QY 307 GGGAGCGCGGAGGTGCGACGAGCAAGCCATATACGCGCGCGCGCGCGCGCGCGCGCG 366
Db 102 AlaAlaAlaAlaAlaGlyGlnSerLysProLysSerLeuValLysProLysSer 121
QY 367 TACATCGCGCTATCGCCATCGCGACTCGCGACTCGCGCGCGCGCGCGCGCGCGCGCG 426
Db 122 TyrIleAlaLeuIleThrMetAlaLeuGlnSerProGlnLysLeuLeuThrLeuSer 141
QY 427 GAGATCAACGAGTACCTCATGGCAAGTTCCTCTTTTCCCGCGCGAGTACACGCGCTGG 486
Db 142 GlyIleCysGluPheIleSerAsnArgPheProTyrTyrArgGluLysPheProAlaTrp 161
QY 487 CGCAACTCCGTCGCGCCACAACTTTCGCTCAACGACTCTCTGCTCAAGTGTGCTGCGCGAC 546
Db 162 GlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPheValLysIleProArgGlu 181
QY 547 CCTCGCGCGCTGGGCAAGGACAACTACTGATGCTCAACCCCAACGAGGAGTACAC 606
Db 182 ProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLeuAspProGlnSerGluAspMet 200
QY 607 TTCGCGCGCGGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 201 PheAspAsnGlySerPheLeuArgArgLysArgPheLysArgHisGlnGlnGluHis 220
QY 661 -----GTCCCGCGCGCG 672
Db 221 LeuArgAspGlnThrAlaLeuMetMetGlnGlyPheGlyAlaTyrGlyLeuAlaGlyPro 240
QY 673 GGGCTGCGCGCGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732
Db 241 TyrGlyArgProTyr-----GlyLeuProProGlyAlaTyrProHisProAlaAla 257
QY 733 CCGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 792
Db 257 ----- 257
QY 793 AAGTTCTCCAGCTCCTTCGCCATCGACAGACATCTCGCAAGAGCCCTTCGAGAGCGGTGCG 852
Db 257 ----- 257
```



Db 18 VallysEluArgGlyGluThrAlaAlaGlyAlaGlyValProGlyGluAlaThrGly 37  
QY 316 GAGGGTGCA-----CGCAGCAGCATATACCGCGCCAGCCAGCCCTACTCG 366  
Db 38 ArgGlyAlaGlyArgArgArgProLeuGlnArgGlyProProTyrSer 57  
QY 367 TACATCGCGCTCATCCGCTGCGCATCCGCGACTCGCGCGGGCGCTGACGCTGGCG 426  
Db 58 TyIleAlaLeuIleAlaMetAlaIleAlaHisAlaProGluArgGlyLeuGly 77  
QY 427 GAGATCAACAGTACCTCATGGCAAGTTCCTTTTCGCGGCGAGCTACACGCGCTGG 486  
Db 78 GlyIleTyrIysPheIleThrGluArgPheProPheTyrArgAspAsnProIysIysIrp 97  
QY 487 CGCAACTCCGTCGCGCCACACCTTCCTCAACGACTGCTTCGCTCAAGTGTGCGCGAC 546  
Db 98 GluAsnSerIleArgHisAsnLeuThrLeuAsnAspCysPheLeuIysIleProArgGlu 117  
QY 547 CCTCGCGCGCTGGCGCAGGACAACTACTGATCTCAACCCACGAGCGATACACC 606  
Db 118 AlaGlyArgPro---GlyIysGlyAsnTyrTrpAlaLeuAspProAsnAlaGluAspMet 136  
QY 607 TTCGCGCAGCGGTCTTCCTCCCGCGCGCGCAAGCGCTCAGC----- 648  
Db 137 PheGluSerGlySerPheLeuArgArgGlyArgPheIysArgSerAspLeuSerThr 156  
QY 649 -----CACCGCGCGCGTCCCGCG----- 669  
Db 157 TyrProAlaTyrMetHisAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 176  
QY 670 -----CCGCGCTCGCGCGCGCGAGGAGCGCGCGCTCCCG 705  
Db 177 AlaAlaAlaAlaAlaIlePheProGlyAlaValProAlaAlaArgProProTyrPro 196  
QY 706 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729  
Db 197 GlyAlaValTyrAlaGlyTyrAlaProProSerLeuAlaAlaProProValTyrTyr 216  
QY 730 GCGCGCGCTCG 777  
Db 217 ProAlaAlaSerPro-----GlyProCysArgValPheGlyLeuValProGluArg 233  
QY 778 GCCACCG 837  
Db 234 -ProLeuSerProGluLeuGlyProAlaProSerGlyProGlyGlySerCysAlaPheAl 253  
QY 838 TTCGCGCA-----CGCGTCCCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCG 888  
Db 253 aSerAlaGlyAlaProAlaThrThrThrGlyTyrGlnProAlaGly-----CysThrG 271  
QY 889 GCGG---CCGCGCGCTCG 945  
Db 271 YAlaArgPro-AlaAsnProSerAlaTyrAlaAlaAlaTyrAlaGlyProAspGlyAla 291  
QY 946 TG---CAGGCGCTGCTCGCGCTCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1002  
Db 291 yrProGlnGlyAlaGlySerAlaIle-----PheAlaAlaAlaGlyA 305  
QY 1003 GCGCGCGAGGCGGAGGTGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062  
Db 305 rGLeuAlaGly-----ProAlaSerProAlaGlyGlySerSerGlyG 320  
QY 1063 GC-----GGCGCGCGCG 1074  
Db 320 lYValGluThrValAspPheTyrGlyArgThrSerProGlyGlnPheGlyAlaLeuG 340  
QY 1075 AAGCACCTCCGAGCG 1134  
Db 340 lYAlaCysTyrAsnProGlyGlyGln----- 348  
QY 1135 GCAGCCCTGCAGCGCGCGCTTAGTCCGCGCTCTCG-----CCGCGACCTGCTCGTAC 1185

Db 349 -----LeuGlyGlyAlaSerAlaGlyAlaTyrHisAlaArgHisAlaAlaAlaTyrP 366  
QY 1186 CCGCTGGA 1193  
Db 366 roGlyGly 368  
RESULT 14  
FXF2\_HUMAN STANDARD; PRT; 444 AA.  
ID FXF2\_HUMAN Q9UQ85;  
AC Q12947; Q9UQ85;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Forkhead box protein F2 (Forkhead-related protein FKHL6) (Forkhead-related transcription factor 2) (FREAC-2) (Forkhead-related activator-2).  
DE FOXF2 OR FKHL6 OR FREAC2.  
GN FOXF2  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99017977; PubMed=9799607;  
RA Blixt A., Mahlapuu M., Bjursell C., Darnfors C., Johannesson T., Enerbaeck S., Carlsson P.;  
RT "The two-exon gene of the human forkhead transcription factor FREAC-2 (FKHL6) is located at 6p25.3";  
RL Genomics 53:387-390(1998).  
RN [2]  
RP SEQUENCE OF 37-444 FROM N.A.  
RX MEDLINE=96224034; PubMed=8626802;  
RA Hellqvist M., Mahlapuu M., Samuelsson L., Enerbaeck S., Carlsson P.;  
RT "Differential activation of lung-specific genes by two forkhead proteins, FREAC-1 and FREAC-2";  
RL J. Biol. Chem. 271:4482-4490(1996).  
RN [3]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=95045392; PubMed=7957066;  
RA Pietrou S., Hellqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;  
RT "Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending";  
RL EMBO J. 13:5002-5012(1994).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=98389768; PubMed=9722567;  
RA Hellqvist M., Mahlapuu M., Blixt A., Enerbaeck S., Carlsson P.;  
RT "The human forkhead protein FREAC-2 contains two functionally redundant activation domains and interacts with TBP and TFIIB";  
RL J. Biol. Chem. 273:23335-23343(1998).  
CC -!- FUNCTION: Probable transcription activator for a number of lung-specific genes.  
CC -!- SUBUNIT: Interacts with the transcription factors TBP and TFIIB.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- TISSUE SPECIFICITY: Lung and placenta.  
CC -!- DOMAIN: Two activation domains, AD1 and AD2, C-terminal of (and distinct from) the forkhead domains are necessary for transcriptional activation.  
CC -!- SIMILARITY: Contains 1 fork-head domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: AF084939; AAD19875.1;  
CC EMBL: AF084938; AAD19875.1; JOINED.  
CC EMBL: UI3220; AAC32226.1;  
CC PIR: T09474; T09474.



```

RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
RN [3]
RP SEQUENCE OF 71-187 FROM N.A.
RX MEDLINE=93387221; PubMed=8375339;
RA Sasaki H., Hogan B.L.;
RT "Differential expression of multiple fork head related genes during
RT gastrulation and axial pattern formation in the mouse embryo.";
RL SUBCELLULAR LOCATION: Nuclear.
CC [-] TISSUE SPECIFICITY: Expressed in many embryonic tissues, including
CC prechondrogenic mesenchyme, periocular mesenchyme, meninges,
CC endothelial cells and kidney.
CC [-] DEVELOPMENTAL STAGE: Expressed during embryogenesis.
CC [-] SIMILARITY: Contains 1 fork-head domain.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF045017; AAC24209.1; -
DR EMBL; L10406; RAA03159.1; -
DR EMBL; X71939; CAA50741.1; -
DR PIR; I49674; I49674.
DR HSP; Q63245; 2HFH.
DR TRANSFAC; T02426; -
DR MGD; MGI:1347466; Foxcl.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FOXHEAD.
DR PRODOM; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK HEAD 1; 1.
DR PROSITE; PS00658; FORK HEAD 2; 1.
DR PROSITE; PS00039; FORK HEAD 3; 1.
DR DNA-binding; Nuclear protein; Transcription regulation.
KW DOMAIN 28 33 POLY-ALA.
FT DOMAIN 77 168
FT DNA_BIND 169 173 FORK-HEAD.
FT DOMAIN 194 197 POLY-ARG.
FT DOMAIN 264 274 POLY-PRO.
FT DOMAIN 375 386 POLY-SER.
FT DOMAIN 444 451 POLY-GLY.
FT DOMAIN 453 456 POLY-SER.
FT DOMAIN 486 496 POLY-GLY.
FT DOMAIN 180 186 VKDKEEK -> KKEITFI (IN REF. 3).
FT CONFLICT 180 186
SQ SEQUENCE 553 AA; 56953 MW; 3CDD12F69CA4F217 CRC64;

Alignment Scores:
Pred. No.: 4.4e-11 Length: 553
Score: 430.50 Matches: 154
Percent Similarity: 43.14% Conservative: 44
Best Local Similarity: 33.55% Mismatches: 119
Query Match: 19.67% Indels: 142
DB: 1 Gaps: 23

US-10-087-080-31 (1-1209) x FXCL_MOUSE (1-553)
Qy 115 TCCTGGGCTCAGATGGGACTGGCGCGCCAGCGCTCCGCGGCGCGCGCGCGAGAT 174
Db 6 SerValSerProAsnSerLeuGlyValValProTyrLeuGly 20
Qy 175 ACGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210
Db 21 -----GlyGluGlnSerTyrArgAlaAlaAlaAlaAlaAlaGlyGly 36
Qy 211 CCGGGCGGAGGAGGAGGATCCCGGCGAGCGAGCTGCTCCAGCGGTGGCGGCGGCG 270
Db 37 -----TyrThrAlaMetPro-----AlaProMetSerValTyrSerHisProAla 51

271 GAGCCCGGGCGCGCGGGGCCAGCGCGCGGGCGCGCGGGAGCGCGGAGGTGCACGAGC 330
52 HisAla-----GluGlnTyrProGlySerMetAlaAlaTyrGly----- 65
331 AAGCCATATACGCGCGCGCC-----AAGCCCTCTACTCTGATATC 372
56 ---ProTyrThrProGlnProGlnProGlnProGlnProGlnProGlnProGlnPro 84
373 CGCTCATCGCATGCGCATCCGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGAGATC 432
85 AlaLeuIleThrMetAlaIleGlnAsnAlaProAspLysLysIleThrLeuAsnGlyLe 104
433 AACGAGTACCTCATGGGCAAGTCCCTTTTCGCGCGCAGCTACACGGGTGCGCAAC 492
105 TyrGlnPheIleMetAspArgPheProPheIleArgAspAsnLysGlnGlyTyrGlnAsn 124
493 TCCTGGCGCCCAACCTTTCCTCAACGACTGCTCTCTCAAGTGTCTGCGCGAGCCCTCG 552
125 SerIleArgHisAsnLeuSerLeuAsnGluCysPheValLysValProArgAspLys 144
553 CGGCGCTGGCGCAGGACAACACTGATGCTCAACCCCAACACAGAGTACACCTTGGCC 612
145 LysPro---GlyTyrGlySerTyrTrpThrLeuAspProAspSerTyrAsnMetPheGlu 163
613 GACGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
164 AsnGlySerPheLeuArgArgArgArgPheLysLysLysAspAlaValLysAspLys 183
646 -----AGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
184 GluGluLysGlyArgLeuHisLeuGlnGluProProProGlnAlaGlyArgGlnPro 203
661 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 720
204 AlaProAlaPro-----ProGluGlnAlaGluGly-----SerAlaPro-GlyPr 218
721 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 759
218 GlnProProProValArgIleGlnAspIleLysThrGluAsnGlyThrCysProSerPr 238
760 GCCCGCCAGGAGGAGC---GCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 816
238 oProGlnProLeuSerProAlaAlaAlaLeuGlySerGlySerAlaAlaThrValPro 258
817 GACAGCATCTCGCAAGCCCTTCGCGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGAG 876
258 sileGluSerProAspSerSerSerSerSerSerSerSerSerSerSerSerSerProGlySe 278
877 ACCTTCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 918
278 rleuProSerAlaArgProLeuSerLeuAspAlaAlaGluProAlaPro-----ProPr 296
919 TTCGCGCGCGCTCCTCC----- 935
296 GlnPro-AlaProProHisSerGlnGlyPheSerValAspAsnIleMetThrS 316
936 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 990
316 erLeuArgGlySerProGlnGlySerAlaAlaGluLeu----- 328
991 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1044
329 --GlySerGlyLeuLeuAlaSerAlaAlaAlaSerArgAlaGlyIleAlaProAl 348
1045 CTGCG-----ACCTCTCC----- 1058
348 euAlaLeuGlyAlaTyrSerProGlnSerSerLeuTyrSerProCysSerGlnS 368
1059 -----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
368 erSerSerAlaGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlySers 388
1081 CTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137

```

```

Db      388 e:SerAlaAlaGly-ThrGlyGlyAlaAlaThrTyHisCysAsnLeuGln---AlaMet 406
QY      1138 GCCCTGCAGCGCGCTTACTCCGNCGTCTCTGCCCGCACCTGCTGTAACCCG 1188
Db      407 SerLeuTyAlaAlaGlyGluArg-----GlyGlyHisLeuGlnGlyPro 421

```

Search completed: March 29, 2004, 12:01:15  
 Job time : 30.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 29, 2004, 11:56:36 ; Search time 67 seconds  
(without alignments)

11386.911 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2189  
Sequence: 1 atgaagtggagggttcgt.....tgagacgcctcctagcttga 1209

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlsp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10087080/runat\_29032004\_114526\_6542/app\_query.fasta\_1.1351  
-DB=SPTREMBL\_25 -QFMT=fastan -SUFFIX=n2p.rspt -MINWATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-NO US10087080@cgn\_1\_86@runat\_29032004\_114526\_6542 -NCPU=6 -ICPU=3  
-DEV WMAP -LARGSEQQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mbc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1597	73.0	400	11 Q9JJ18	Q9jj18 mus musculus

2	1534	70.1	400	11 Q9JLN7	Q9jln7 rattus ratt
3	533.5	24.4	492	11 Q35392	Q35392 mus musculus
4	506	23.1	443	13 P79770	P79770 gallus gall
5	467	21.3	371	11 Q8R210	Q8r210 mus musculus
6	449.5	20.5	370	11 Q08771	Q08771 rattus norv
7	431.5	19.7	553	4 Q86UP7	Q86up7 homo sapien
8	430.5	19.7	553	11 Q9QWR9	Q9qwr9 mus musculus
9	425.5	19.4	446	11 Q54743	Q54743 mus musculus
10	411.5	18.8	329	11 Q8BIK9	Q8bik9 mus musculus
11	409.5	18.7	377	6 Q8MIP2	Q8mip2 capra hircu
12	405.5	18.5	375	11 Q8K3X9	Q8k3x9 mus musculus
13	402.5	18.4	528	13 Q93440	Q93440 gallus gall
14	402	18.4	371	13 Q9DEN4	Q9den4 xenopus lae
15	394.5	18.0	330	4 Q96D28	Q96d28 homo sapien
16	393	18.0	371	13 Q9DEN3	Q9den3 xenopus lae
17	384	17.5	494	11 Q8C694	Q8c694 mus musculus
18	379.5	17.3	402	5 Q8MUS6	Q8mus6 brachyosteo
19	379.5	17.3	445	13 P79771	P79771 gallus gall
20	370	16.9	468	13 Q13017	Q13017 gallus gall
21	366	16.7	371	13 Q73782	Q73782 brachydanio
22	365.5	16.7	442	5 Q93133	Q93133 brachyosteo
23	365	16.7	383	13 Q7T1C2	Q7t1c2 brachydanio
24	360	16.4	329	11 Q8CSN7	Q8csn7 mus musculus
25	359	16.4	353	13 Q73785	Q73785 brachydanio
26	359	16.4	492	13 Q9PVZ3	Q9pvz3 xenopus lae
27	358	16.4	383	13 Q7T1C0	Q7t1c0 brachydanio
28	352.5	16.1	456	5 Q8T4G1	Q8t4g1 drosophila
29	352.5	16.1	506	5 Q95YK7	Q95yk7 ciona savig
30	352	16.1	346	13 Q9OWN4	Q9own4 xenopus lae
31	352	16.1	492	13 Q9YHB2	Q9yhb2 xenopus lae
32	349	15.9	348	13 Q8JIE3	Q8jie3 oreochromis
33	348.5	15.9	352	13 Q9PRJ8	Q9prj8 xenopus lae
34	348.5	15.9	353	13 Q9PT68	Q9pt68 xenopus lae
35	348.5	15.9	353	13 Q7T1C1	Q7t1c1 brachydanio
36	347	15.9	345	11 Q922I5	Q922i5 mus musculus
37	347	15.9	372	11 Q9D299	Q9d299 mus musculus
38	347	15.9	583	5 Q9GUX6	Q9gux6 ciona savig
39	346.5	15.8	381	13 Q8JIT5	Q8jit5 xenopus lae
40	345.5	15.8	411	13 Q90868	Q90868 gallus gall
41	345.5	15.8	444	13 Q98TD2	Q98td2 gallus gall
42	344.5	15.7	438	13 Q9PWP8	Q9pwp8 gallus gall
43	344	15.7	367	13 Q91905	Q91905 xenopus lae
44	341.5	15.6	372	13 Q9W707	Q9w707 xenopus lae
45	341.5	15.6	373	13 Q9W706	Q9w706 xenopus lae

## ALIGNMENTS

### RESULT 1

Q9JJ18	Q9JJ18	PRELIMINARY;	PRT;	400 AA.
ID	Q9JJ18			
AC	Q9JJ18			
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-JUN-2003 (Tremblrel. 24, Last annotation update)			
DE	HFH1 (Forkhead box Q1).			
GN	FOXQ1 OR HFH1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129;			
RC	MEDLINE=21207087; PubMed=11309849;			
RA	Hong H.-K.; Noveroske J.K.; Headon D.J.; Liu T.; Sy M.S.;			
RA	Justice M.J.; Chakravarti A.;			
RT	"The winged helix/forkhead transcription factor Foxq1 regulates			
RT	differentiation of hair in satin mice.";			
RL	Genesis 29:163-171(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=FVB/N-3; TISSUE=Breast tumor;			
RC				

```

RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N-3; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154426; AAF74524.1; -
DR EMBL; BC047155; AAH47155.1; -
DR HSSP; Q63245; 2HFH.
DR MGD; MG1:1298228; Foxq1.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PD00053; FORKHEAD.
DR PRODOM; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
SQ SEQUENCE 400 AA; 41368 MW; F286AEPF81D4759B CRC64;

Alignment Scores:
Pred. No.: 2,15e-73 Length: 400
Score: 1597.00 Matches: 320
Percent Similarity: 82.27% Conservative: 14
Best Local Similarity: 78.82% Mismatches: 62
Query Match: 72.96% Indels: 10
DB: 6 Gaps: 6

US-10-087-080-31 (1-1209) x Q9JUL8 (1-400)
QY 1 ATGAAGTTGGAGGTGTTCTGCTCCCTCGCGCGGCCCAAGCAGGCGCAGTGACCTG 60
Db 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysMetGlySerAspLeu 20
QY 61 GAGGCGCGGCGGACGACGCGCGCTCCCGCTGTCGCGCGGCGGAGCAGCACTCCCTG 120
Db 21 GluGlyAlaGlySerAspValProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
QY 121 GGCTCAGATGGGACTCGCGGCC--AAGCCGTCCGCGGCGCGCGCGCAGAGATACG 177
Db 41 GlySerAspGlyAspCysAlaAlaAsnSerProAlaAlaGlySerGlyAlaGlyAspLeu 60
QY 178 CAG---GGCAGCGCGCAACAGAGTGGGAGGCGGCGCGCGCGCGGAGGCGGATCCCG 234
Db 61 GluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 80
QY 235 GCACGAGCTGCTGACGGGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 294
Db 81 AlaThrAspAspSerArg-----ThrGlnAlaSerAlaAlaGlyProCys 95

```

```

QY 295 GCGGGCGCGCGCGGAGCGGCGGAGGCTGCACGAGCAAGCATATACGGCGCGGCCAAG 354
Db 96 AlaGlyGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 115
QY 355 CCCCCCTACTCGTACATCGCGCTCATCGCATCGCATCGCATCGCATCGCATCGCATCG 414
Db 116 ProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleAlaArgAspSerAlaGlyArg 135
QY 415 TTGACGCTGGCGGAGATCAACGAGTACCTCATCGGGAAGTCCCTTTTCCCGGGCAGC 474
Db 136 LeuThrLeuAlaGluIleAenGluTyrLeuMeGlyLysPheProPheAspArgGlySer 155
QY 475 TACACGGCTGGCGCAACTCCGTCGCGCCACACACTTTCGCTCAACGACTCTTCGTCG 534
Db 156 TyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheVallys 175
QY 535 GTGCTGGCGACCCCTCGCGCCCTGGGCAAGAGCAACTACTGAGTGCCTCAACCCCAAC 594
Db 176 ValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsn 195
QY 595 AGCGAGTACACCTTCGCGCGGCGGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 654
Db 196 SerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLysLeuSerHisArg 215
QY 655 GCGCGGCTCCCGCGCGCGCGCGCTGCGCCCGGAGAGCGCGCGCGCTCCCGCGCG--GCC 711
Db 216 ThrThrValSerAlaSerGlyLeuArgProGluGluAlaProGlyProAlaGlyThr 235
QY 712 CCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 771
Db 236 ProGlnProAlaProAlaAlaArgSerSerProIleAlaArgSerProAlaAlaArgGlnGlu 255
QY 772 GAGCGCGCGCGCGCGCGCGCGCGAGTTCTCAGCTCTCTCCGATCGACAGCATCTCGCGC 831
Db 256 GluArgSerSerProAlaSerLysPheSerSerSerPheAlaIleAspSerIleLeuSer 275
QY 832 AAGCCCTTCGCGACGCTCGCTCAGGCGACAGCGCGCGCGCGCGCGCGCTTCAGTGGCGC 891
Db 276 LysProPheArgSerArgArgAspGlyAspSerAlaLeuGlyValGlnLeuProTyrGly 295
QY 892 GCGCGCGCTGCGCGCGCGCGCTGCGCGCTTCCCGCGCTTCCCGCGCGCGCGCGCGCGCG 951
Db 296 AlaAlaProCysProProLeuArgAlaTyrProAlaLeuLeuProAlaAlaProGlyGly 315
QY 952 GCCCTGCTGCGCGCTCGCGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1011
Db 316 AlaLeuLeuProLeuCysAlaTyrGlyAlaSerGluProThrLeuLeuAlaSerArgGly 335
QY 1012 GCGAGGTGCCACCGCGCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1071
Db 336 ThrGluValGlnProAlaAla---ProLeuLeuLeuAlaProLeuSerThrAlaAlaPro 354
QY 1072 GCCAAGCCACTCCGAGGCGCG--GCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTG 1128
Db 355 AlaLysProPheArgGlyProGluThrAlaGlyAlaAlaHisLeuTyrCysProLeuArg 374
QY 1129 CTGCGCGCGCGCGCGCGCGCGCTTAGTCGCGCGCTCTGCGCGCGCGCGCGCGCGCGCG 1188
Db 375 LeuProThrAlaLeuGlnAlaAlaAlaAlaCysGlyProGlyProHisLeuSerTyrPro 394
QY 1189 GTGAGAGCTCTTAGCT 1206
Db 395 ValGluThrLeuLeuAla 400

RESULT 2
Q9JUL7 PRELIMINARY; PRT; 400 AA.
ID Q9JUL7;
AC 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Winged helix/forkhead transcription factor HFH1.
GN HFH1.

```





DR Pfam: PF00250; Fork head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR ProDom; PD000425; TF\_Fork\_head; 1.  
 DR SMART; SM00339; FH; 1.  
 DR PROSITE; PS00657; FORK HEAD 1; 1.  
 DR PROSITE; PS00658; FORK HEAD 2; 1.  
 DR PROSITE; PS00039; FORK HEAD 3; 1.  
 SQ SEQUENCE 443 AA; 46429 MW; 409563037EA5911F CRC64;

Alignment Scores:  
 Pred. No.: 3,648-18 Length: 443  
 Score: 506.00 Matches: 160  
 Percent Similarity: 43.18% Conservative: 33  
 Best Local Similarity: 35.79% Mismatches: 154  
 Query Match: 23.12% Indels: 100  
 DB: 13 Gaps: 20

US-10-087-080-31 (1-1209) x P79770 (1-443)

QY 64 GCGCGCGCGCGC-----ACGACGCGCGCTCCCGCTGCGCG 102  
 DB 7 GlyGlyGlyGlyCysGlyIleMetSerGluArgSerProGluGluProLeuSerGlu 26  
 QY 103 GCGGAGAC-----CACTCCCTGGCTCAGATGGGACTGCGCGCAAGCCGCC 153  
 DB 27 ValGluAspAlaAspIleAspValValGlyProGlnAsp---GlyAlaLysTyrSer 45  
 QY 154 GCGGCGCGCGCGCGCAGATACGAGGCGCGCAACAGTGCAGGAGCGCGCGCG 213  
 DB 46 GluAspGluGluAspAspAspGluGluAspGluGluGlyGlyPro 65  
 QY 214 -----GGCGGAGGAGCG 228  
 DB 66 TrpGlySerProAlaAlaAspGlyGlyProProSerAlaHisGlyGlyValProGluArg 85  
 QY 229 ATCCGCGCAGAGCTCTCAGCGTGTGGGAGGCGCGGAGCGCGCGCGCGCG 288  
 DB 86 LeuSerProAlaGlyAlaArgSerProArgAlaProGlyProArgProGlyLysArgAla 105  
 QY 289 CCAGCG 348  
 DB 106 AlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyLysAsn---ProLeuVal----- 122  
 QY 349 CCCAGCGCCCTACTCGTACATCGCTCATCGCATGCGCATCGCGACTCGCGCGCG 408  
 DB 123 ---LysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleLeuGlnSerProLys 141  
 QY 409 GCGCGCTTACGCTGCGGAGATCAACGAGTACCTCATGGCAAGTTCCTCTTCGCG 468  
 DB 142 LysArgLeuThrLeuSerGluIleCysGluPheIleSerGlyArgPheProTyrTyrArg 161  
 QY 469 GCGAGTACACGGCTGCGGCACTCCGTCGCGCAACCTTCGCTCAAGACTGCTTC 528  
 DB 162 GluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPhe 181  
 QY 529 GTCAAGTCTGCGGACCCCTCGCGCGCTCGGCGCGCGCGCGCGCGCGCGCGCG 588  
 DB 182 ValLysIleProArgGluProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLeuAsp 200  
 QY 589 CCCACAGCGAGTACACCTTCCGCGAGCGGCTCTCCGCGCGCGCGCGCGCGCGCG 648  
 DB 201 ProGluSerAlaAspMetPheAspAsnGlySerPheLeuArgArgArgLysArgPheLys 220  
 QY 649 ---CACCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 684  
 DB 221 ArgHisGluGluProAspProProHisProGluLeuLeuLeuArgAlaGlyAlaArgSer 240  
 QY 685 GAGGAGCGCGCG----- 696  
 DB 241 ArgArgLeuProAlaArgLeuArgTyrGlyProTyrGlyTyrAsnTyrGlyLeuGlnLeu 260  
 QY 697 ---GGCTCCCG 753

DB 261 GlnGlyLeuProProGly---ProProAla-ProProProProProArg----- 275  
 QY 754 TCG 807  
 DB 276 -ArgArgArgLeuArgGlyAlaPheProPheSerAlaProHisCysProLeuProValGln 295  
 QY 808 TTCGCCATCGACAGCATCTGCGCAAGCCCTTCCGAGCGCGTCTCAGGGACACGGCC 867  
 DB 295 yProProSerAlaAlaSerValPheSerAlaAlaSerGlyLeuProSer----- 311  
 QY 868 CCGGCGCGCGCGCTTCAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 927  
 DB 312 -----PheLeuGlyGlyLeuLeuAsnCysArgLysSerPheTyrHisProGln 327  
 QY 928 CTCCTCCCG 987  
 DB 327 nLeuSerProThr-AlaLeu-----ProAlaAlaLeuLeuGlnThrLeuLysProAsp 345  
 QY 988 CCGGCG 1037  
 DB 345 roThrProAlaGlyThrGlyGlyThrAlaAlaAlaThrAsnProSerArgProSerP 365  
 QY 1038 -----CCTCTGCTTGCACCTCTCCCGCGCG 1062  
 DB 365 heSerIleAspAsnIleIleGlyAlaValProProProProSerThrAsnProSerA 385  
 QY 1063 GCG 1112  
 DB 385 laAlaPro-----AlaAlaProTyrProSerGlyGlnAlaGlyProProAlaGlnLeu 403  
 QY 1113 ---GTACTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1158  
 DB 403 euAlaValLeuSerProAlaLeuAlaProSerGlnProHisGlyGlyLeuAlaHisGlu 423  
 QY 1159 CGNCGTCTCGCGCGCG 1175  
 DB 423 roLeuLeuGlnProAla 428

RESULT 5  
 Q8R210 PRELIMINARY; PRT; 371 AA.  
 ID Q8R210 AC  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Forkhead box E1.  
 GN FOXE1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RA MEDLINE=22190884; PubMed=12203737;  
 RX Dathan N., Parlato R., Rosica A., De Felice M., Di Lauro R.;  
 RT "Distribution of the ttf2/foxe1 gene product is consistent with an  
 RT important role in the development of foregut endoderm, palate and  
 RT hair";  
 RL Dev. Dyn. 224:450-456(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RA Parlato R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ457047; CAD29716.2; .  
 SQ SEQUENCE 371 AA; 37767 MW; 8AB92160F2DDFAFA CRC64;

Alignment Scores:  
 Pred. No.: 3,47e-16 Length: 371  
 Score: 467.00 Matches: 150  
 Percent Similarity: 46.44% Conservative: 26  
 Best Local Similarity: 39.58% Mismatches: 131



```

Db      81  TyLysPheIleThrGluArgPheProPheTyArgAspAsnProLysLeuTyrGlnAsn 100
QY      493  TCGGTGGCGCAACCTTTCGGTCAACGACTCTCGTCAAGGTCTCGGCGACCCCTCG 552
Db      101  SerIleArgHisLeuThrLeuAsnAspCysPheLeuLysIleProArgGluAlaGly 120
QY      553  CGGCTTCGGGCAAGCAACTACTGATGCTCAACCCCAACAGGAGTACACCTTCGCC 612
Db      121  ArgPro---GlyLysGlyAsnTyrTrpAlaLeuAspProAsnAlaGluAspMetPheGlu 139
QY      613  GACGGGTCTTCGGCGCGCGCGCAAGCGCTCAGC----- 648
Db      140  SerGlySerPheLeuArgArgArgArgPheLysArgSerAspLeuSerThrTyrPro 159
QY      649  -----CACCGCGCGCGTCCCGCG-----CCCGGG 675
Db      160  AlaTyrMetHisAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaPheProGly 179
QY      676  CTGGCGCGCGAGGCGCGCGCTCCGC----- 705
Db      180  AlaValProAlaAlaArgProAlaTyrProGlyAlaValTyrAlaGlyTyrAlaProPro 199
QY      706  ---GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 762
Db      200  LeuAlaAlaProProValTyrTyrProAlaAlaSerPro-----GlyProCys 216
QY      763  CGC-----CAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810
Db      217  ArgValPheGlyLeuValProGluArg-ProLeuSerProAspLeuGlyProAlaProSe 236
QY      811  GCATCGCAGCATCTCGCGCAAGCCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 870
Db      236  rAlaAlaGlyGlySerCysAlaPheAlaAlaAlaAlaGlyAlaProGlyThr-GlySerP 256
QY      871  GGGACGACGCT---TCAGTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 924
Db      256  heGlnProAlaValCysThrGlyAlaArgProValAsnProAlaAlaTyrAlaAlaAla 276
QY      925  GGTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 981
Db      276  yAlaAlaGlyProAspGlyAlaTyrProGlnGlyAlaSerSerAlaLeu----- 291
QY      982  GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1041
Db      292  --PheAlaAlaAlaAlaGlyArgLeuAlaGly----- 1082
QY      1042  CTGCTTGCACTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1082
Db      303  hrSerProThrAlaGlyGlyGlySerGlyGlyValGlnAlaThrValAspPheTyrGlyA 323
QY      1083  -----CCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1134
Db      323  rgThrSerProGlyGlnPheGlyAlaAlaLeuGlyProCysTyrAsnProSerGlyGlnL 343
QY      1135  GCAGCCCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1190
Db      343  euGlyAlaGlyGlyGlyAlaTyrHisSerArgHisAlaThrAlaTyrProGly 361
RESULT 7
Q86UP7  PRELIMINARY; PRT; 553 AA.
AC      Q86UP7;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Forkhead winged/helix transcription factor mutant 2.
GN      FOXCI.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Chiroptera; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]

```

```

SEQUENCE FROM N.A.
RA      Konatireddy S., Chakrabarti S., Mandal A.K., Reddy A.B., Sampath S.,
RA      Panicker S.G., Balasubramanian D.;
RT      "Human FOXCI gene.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY228704; AAP15181.1; -.
DR      GO; GO:0005634; C:nucleus; IEA.
DR      GO; GO:0003700; T:transcription factor activity; IEA.
DR      GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR      InterPro; IPR001766; TF_Fork_head.
DR      Pfam; PF00250; Fork_head; 1.
DR      PRINTS; PR00053; FORKHEAD.
DR      ProDom; PD000425; TF_Fork_head; 1.
DR      SMART; SM00339; FH_1.
DR      PROSITE; PS00657; FORK_HEAD_1; 1.
DR      PROSITE; PS00658; FORK_HEAD_2; 1.
DR      PROSITE; PS00039; FORK_HEAD_3; 1.
SQ      SEQUENCE 553 AA; 56784 MW; B2B7017111CE5125 CRC64;

Alignment Scores:
Pred. No.:      2,11e-14      Length:      553
Score:          431.50      Matches:    146
Percent Similarity: 42.05%      Conservative: 39
Best Local Similarity: 33.18%      Mismatches: 123
Query Match:      19.71%      Indels:     132
DB:              4          Gaps:       18

US-10-087-080-31 (1-1209) x Q86UP7 (1-553)
QY      214  GCGCGGAGAGAGCGGATCCCGGCGAGCTGCTGCGAGCGTGTGGCGGAGGCGCGGAG 273
Db      20  GlyGlyGlnSerTyrTyrArgAlaAlaAlaAla-----AlaGlyGlyGlyTyr 37
QY      274  GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
Db      38  ThrAlaMetProAlaProMetSerValTyrSerHisProAlaHisAlaGluGlnTyrPro 57
QY      298  GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 351
Db      58  GlyGlyMetAlaArgAlaTyrGly-----ProTyrThrProGlnProGlnPro 73
QY      352  -----AAGCCCGCTACTCGTACATCGGCTCATCGGCTCATCGGCTCATCGGCT 399
Db      74  LysAspMetValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleGlnAsn 93
QY      400  TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459
Db      94  AlaProAspLysLysIleThrLeuAsnGlyIleTyrGlnPheIleMetAspArgPhePro 113
QY      460  TTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 519
Db      114  PheTyrArgAspAsnLysGlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsn 133
QY      520  GACTGCTTCGTCAGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579
Db      134  GlucySphValLysValProArgAspAspLysLysPro---GlyLysGlySerTyrTyr 152
QY      580  ATGCTCAACCCCAACACGCGAGTACACCTTCGCGCGCGCGCGCGCGCGCGCGCG 639
Db      153  ThrLeuAspProAspSerTyrAsnLysPheGluAsnGlySerPheLeuArgArgArg 172
QY      640  CGCCTCAGC-----CACCGCGCG 657
Db      173  ArgPheLysLysLysAspAlaValLysAspLysGluLysAspArgLeuHisLeuLys 192
QY      658  CCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708
Db      193  GluProProProGlyArgGlnProProProAlaProProGluGlnAlaAspGlyAsn 212
QY      709  GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
Db      213  AlaPro-GlyProGlnProProProValArgIleGlnAspIleLysThrGluAsnGlyTh 232

```

```

QY 748 ATGCGCTCCCGCCCGCCGAGGAGGAGC---GCGCCAGCCCGCCGCGGCAAGTTCCTCCAGC 804
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 rCysProSerProGlnProLeuSerProAlaAlaAlaLeuGlySerGlySerAlaAl 252
QY 805 TCCTTCGCGCATCGACAGCATCTCGCAAGCCTTCGCGAGCGCGTCCAGGACACG 864
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 alalavalProLysleuSerProAspSerSerSerSerSerSerSerSerGlySerSe 272
QY 865 GCGCCCGGGA-----CGACGCTTCAGT-----GGGCGGCC 894
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 rProProGlySerLeuProSerAlaArgProLeuSerLeuAspGlyAlaAspSerAlaPr 292
QY 895 GCGCCCTGCGCGCGTCCGCGGTTCGCGGCTCTCC----- 935
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 oProPro-ProAla-----ProSerAlaProProProHisHisSerGlnGlyP 308
QY 936 -----CGCGCGCCCTCGAGGCGCCCTGCGCGCTGCTGCGCGTC 966
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
308 heSerValAspAsnIleMetThrSerLeuArgGlySerProGlnSerAlaAlaGluL 328
QY 967 TGGCGGTACGGCGCGGAGCGCGCGTGGCGCGCGCGAGCGCGAGTGCACCG 1026
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
328 eu-----SerSerGlyLeuLeuAlaSerAlaAlaAsers 340
QY 1027 AC CGCGCC-----GCCCTCTCT----- 1043
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
340 erArgAlaGlyIleAlaProProLeuAlaLeuGlyAlaTySerProGlyGlnSerSerL 360
QY 1044 -----GCTTGCACCTCTCCGCGCGCGCGCGCGCGCGCC 1074
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
360 euTySerSerProCysSerGlnThrSerSerAlaGlySerSerGlyGlyGlyGlyG 380
QY 1075 AAGCCACTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1128
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
380 lyAlaGlyAlaAlaGlyGlyAlaGlyGlyAlaGlyThyTyHisCysAsnLeuGlnAla 399
QY 1129 -----CTGCCCGCAGCGCTG 1143
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 MetSerLeuTyAlaAlaGlyGluArgGlyGlyHisLeuGlnGlyAlaProGlyGlyAla 419
QY 1144 CAGCGCGCTTACCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1197
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
420 GlyGlySerAlaValAspAsnProLeuProAspTySerLeuProValThr 437

RESULT 8
Q9QWR9 PRELIMINARY; PRT; 553 AA.
AC Q9QWR9
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE FKHI/MF1 protein (Hypothetical protein).
GN FOXC1 OR FKHI/MF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=98440371; PubMed=9767123;
RC TrEMBL=2238257;
RX Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

```

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellaro N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.B.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL [3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RP Strauberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ232398; CAAL1239.1; -
DR EMBL; EC052011; AAHS2011.1; -
DR HSSP; Q63245; 2HFF.
DR MGD; MGI:1347466; Foxcl.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00339; FORK_HEAD_3; 1.
KW Hypothetical protein.
SQ SEQUENCE 553 AA; 56939 MW; 3CDB14F69AA2F217 CRC64;

Alignment Scores:
Pred. No.: 2,38e-14 Length: 553
Score: 430.50 Matches: 154
Percent Similarity: 43.14% Conservative: 44
Best Local Similarity: 33.55% Mismatches: 119
Query Match: 19.67% Indels: 142
DB: 23

US-10-087-080-31 (1-1209) x Q9QWR9 (1-553)
QY 115 TCCTGGGCTCAGATGGGAGTGGCGCGCAAGCGTCCGCGCGCGCGCGCGAGAT 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6 SerValSerSerProAsnSerLeuGlyValValProTyLeuGly----- 20
QY 175 ACGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
21 -----GlyGluGlnSerTyTyArgAlaAlaAlaAlaAlaAlaGlyGly 36
QY 211 CCGCGCGCGGAGGAGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 270
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
37 -----TyrThrAlaMetPro-----AlaProMetSerValTySerHisProAla 51
QY 271 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 330
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 HisAla---GluGlnTyProGlySerMetAlaArgAlaTyGly----- 65
QY 331 AAGCCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 ---ProTyThrProGlnProGlnProLysAspMetValTySerProTySerTyLe 84
QY 373 GCGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 AlaLeuIleThrMetAlaIleGlnAsnAlaProAspLysIleThrLeuAsnGlyLe 104

```

```
QY 433 AACGAGTACCTCATGGGCAAGTTCCTCTTTTCGGCGGCACTACACGGGCTGCGCAAC 492
DB 105 TyrGlnPheIleMetAspArgPheProPheTyrArgAspAsnLysGlnGlyTrpGlnAsn 124
QY 493 TCCGTGGCCCAACCTTTTCGTCACACGACTGCTTCGTCAGAGTGTGCGGACCCCTCG 552
DB 125 SerIleArgHisAsnLeuSerLeuAsnGluCysPheValLysValProArgAspAsnLys 144
QY 553 CGGCCTCGGCAAGGACCACTACTGATGTCACACCCACACAGAGTACACCTTCGCC 612
DB 145 LysPro--GlyLysGlySerTyrTrpThrLeuAspProAspSerTyrAsnMetPheGlu 163
QY 613 GACGGGTCTCCGCGCGCGCGCGCGCGCTC----- 645
DB 164 AsnGlySerPheLeuArgArgArgArgPheLysLysAspAlaValLysAspLys 183
QY 646 -----AGCACCGCGCGCGCG 660
DB 184 GluGluLysGlyArgLeuHisLeuGlnGluProProProGlnAlaGlyArgGlnPro 203
QY 661 GTCCCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCG 720
DB 204 AlaProAlaPro-----ProGluGlnAlaGluGly-----SerAlaPro-GlyPr 218
QY 721 GCGCGCGCGCGCGCG-----CGGCTCGCGCGCGCGCGCGCGCTCGCGCGCG 759
DB 218 GlnProProProValArgIleGlnAspIleLysThrGluAsnGlyThrCysProSerPr 238
QY 760 GCCCGCCAGGAGGAGC---GCCCGAGCGCGCGCGCGCGAGTCTCCAGCTCTCCGCA 816
DB 238 oProGlnProLeuSerProAlaAlaLeuGlySerGlySerAlaAlaThrValProLy 258
QY 817 GACAGATCTCGCAAGCCCTTCGCGAGCGCTCGCGCGCGCGCGCGCGCGCGCGAG 876
DB 258 sIleGluSerProAspSerSerSerSerLeuSerSerGlySerSerProProGlySe 278
QY 877 AGCTTTCAGTGGGCGCGCG-----CGCGCTCGCGCGCGCGCGCGCGCGCG 918
DB 278 rLeuProSerAlaArgProLeuSerLeuAspAlaGluProAlaPro-----ProPr 296
QY 919 TTCGCGCGCGCTCTCC----- 935
DB 296 oGlnPro-AlaProProProHisSerGlnGlyPheSerValAspAsnIleMetThr 316
QY 936 -----CGCGCGCGCGCTCGCGCGCGCTCGCGCGCTCGCGCGCGCGCGCG 990
DB 316 erLeuArgGlySerProGlnGlySerAlaAlaGluLeu----- 328
QY 991 GCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1044
DB 329 --GlySerGlyLeuLeuAlaSerAlaAlaAlaSerArgAlaGlyIleAlaProL 348
QY 1045 CTTGC-----ACCTCTCC----- 1058
DB 348 euAlaLeuGlyAlaTyrSerProGlyGlnSerSerLeuTyrSerProCysSerGln 368
QY 1059 -----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
DB 368 erSerSerAlaGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 388
QY 1081 CTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137
DB 388 erSerAlaAlaGly-ThrGlyGlyAlaAlaThrTyrHisCysAsnLeuGln---AlaMet 406
QY 1138 GCCTCTCAGCGCGCGCTTAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1188
DB 407 SerLeuTyrAlaAlaGlyGluArg-----GlyGlyHisLeuGlnGlyPro 421
```

RESULT 9

```
OS4743
ID OS4743 PRELIMINARY; PRT; 446 AA.
AC OS4743;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
```

```
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
LN
GN FOXF2 OR FKX20 OR LUN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Miura N., Kakinuma H., Sato M., Aiba N., Sugiyama T.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 11-446 FROM N.A.
RA Miura N., Kakinuma K., Sato M., Aiba N., Sugiyama T.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y12293; CAA22972.1; -.
DR EMBL; Y12294; CAA22972.1; JOINED.
DR EMBL; Y11148; CAA72035.1; -.
DR HSP; O63245; 2HPH.
DR TRANSFAC; T02514; -.
DR MGD; MGI:1347479; Foxf2.
DR GO; GO:0005634; Cnucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR PRODOM; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
SQ SEQUENCE 446 AA; 46374 MW; 472303F238C016A7 CRC64;
```

## Alignment Scores:

```
Pred. No.: 4, 32e-14 Length: 446
Score: 425.50 Matches: 128
Percent Similarity: 40.81% Conservative: 34
Best Local Similarity: 32.24% Mismatches: 152
Query Match: 19.44% Indels: 83
DB: 11 Gaps: 11
```

US-10-087-080-31 (1-1209) x OS4743 (1-446)

```
QY 76 AGCGACGCGCGCGCTCCCGCTGTCGGCGCGGAGACGACTCCCTGGCTCAGATGGGAC 135
DB 24 SerProAlaProGlyAlaLeuGlnAlaAla----- 33
QY 136 TGGCGGCCCAAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 195
DB 34 LeuMetSerProProProAlaAlaThrLeuGluSerThrSerSerSerSerSer 53
QY 196 AGTGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 255
DB 54 SerAlaSerCysAlaSerSerSerSerSerSerSerSerSerSerSerSerSer 69
QY 256 GTGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 315
DB 70 -----GlyAlaCysLysSerAlaAlaSerSerGlyGlyAlaGlySerGly 86
QY 316 GAGGGTGCACGACGACGACCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
DB 87 GlyThrLysLysAlaThrSerGlyLeuArgProGluLysProTyrSerTyrIle 106
QY 373 GCGCTCATCCCATGGCCATCCCGACTCCGCGCGCGCGCGCGCGCGCGCGCGCG 432
DB 107 AlaLeuIleValMetAlaIleGlnSerSerSerSerSerSerSerSerSerSer 126
QY 433 AACGAGTACCTCATGGGCAAGTTCCTCTTTTCGGCGGCGAGCTACACGGGCTGCGCAAC 492
DB 127 TyrGlnPheLeuGlnAlaArgPheProPheArgGlyAlaTyrGlnGlyTyrLysAsn 146
```

QY 493 TCCTGGCGGCACACCTTTTCCTCAACGACTGCTTCTGCTCAAGGTGCTGGCGACCCCTCG 552  
Db 147 SerValArgHisLeuSerLeuAenGluCysPheIleLeuSerProLysGlyLeuGly 166  
QY 553 CGGCTTGGGCGAGGACAACTACTGATGCTCAACCCCAACAGCAGATACACCTTCGCC 612  
Db 167 ArgPro---GlyLysGlyHisTyrTrpThrIleAspProAlaSerGluPheMetPheGlu 185  
QY 613 GAGCGGTCTTCGCGCCCGCCGCGCAGCGCTCAGCCACCGCGCGCGCTCCCGCGGCC 672  
Db 186 GluGlySerPheArgArgProArgGlyPheArgArgLysCysGlnAlaLeuLysPro 205  
QY 673 -----GGCTG----- 678  
Db 206 MetTyrHisArgValSerGlyLeuGlyPheGlyAlaSerLeuLeuProGlnGlyPhe 225  
QY 679 -----CGGCGCGAGGAGGCCG-----GGC 699  
Db 226 AspPheGlnAlaProProSerAlaProLeuGlyCysHisGlyGlnGlyTyrGlyGly 245  
QY 700 CTCCCGCCCGCCCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 759  
Db 246 LeuAspMetMetProAlaGlyTyrAspThrGlyAlaGlyAla-ProGlyHisAlaHisPr 265  
QY 760 GCCCGCAGGAGGCG 819  
Db 265 oGlnHisLeuHisHisHisValProHisMetSerProAsnProGlySerThrTyrMe 285  
QY 820 AGCATCTTGGCGAGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 879  
Db 285 tAlaSerCysProValProAlaGlyProAlaGlyValGlyAlaAlaGlyGlyGlyGly 305  
QY 880 CTTCAGTGGGCG 939  
Db 305 yGlyGlyGlyAspTyrGlyPro-----AspSerSerSerSerPr 318  
QY 940 GCGCGCTTGGCG 999  
Db 318 oValProSerSerProAlaMetAlaSerAla----- 328  
QY 1000 GGCGCGCGCGAGGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1053  
Db 329 -----IleGluCysHisSerPro-TyrThrSerProAlaAlaHisTrpSerS 344  
QY 1054 CTCCCG 1113  
Db 344 eProGlyAlaSerProTyrLeuLysGlnPro-----ProA 356  
QY 1114 TACTGCCCTTGGCG 1160  
Db 356 laLeuThrProSerSerAsnProAlaAlaSerAlaGlyLeuHisPro 371  
RESULT 10  
Q8BIK9 PRELIMINARY; PRT; 329 AA.  
ID Q8BIK9  
AC Q8BIK9;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Forkhead box protein 11.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=22354683; PubMed=12466853;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).

EMBL; AK045276; BAC32291.1; -.  
PIR; PT0706; PT0706.  
GO; GO:0005634; C:nucleus; IEA.  
GO; GO:0003700; P:transcription factor activity; IEA.  
GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
InterPro; IPR001766; TF\_Fork\_head.  
Fam; F00250; Fork\_head; 1.  
PRINTS; PR00053; FORKHEAD.  
ProDom; PD000425; TF\_Fork\_head; 1.  
SMART; SM00339; FH; 1.  
PROSITE; PS00657; FORK\_HEAD\_1; 1.  
PROSITE; PS00658; FORK\_HEAD\_2; 1.  
PROSITE; PS00039; FORK\_HEAD\_3; 1.  
SQ SEQUENCE 329 AA; 34570 MW; 06BB830798781AB9 CRC64;  
Alignment Scores:  
Pred. No.: 2,26e-13 Length: 329  
Score: 411.50 Matches: 117  
Percent Similarity: 45.03% Conservative: 37  
Best Local Similarity: 34.21% Mismatches: 90  
Query Match: 18.80% Indels: 98  
DB: 11 Gaps: 14  
US-10-087-080-31 (1-1209) x Q8BIK9 (1-329)  
QY 76 AGCGAGCGCGCTCCCGCTGTCGCGCGCGCGAGAC-----GAC 114  
Db 5 ThrGluProAlaProAlaGlnAlaGlyGlyLeuAspMetAlaGlyPheCysAsp 24  
QY 115 TCCCTGGCTCAGATGGGACTGC-----CGGCC 144  
Db 25 SerLeuGlySer-----CysSerValProHisGlyLeuThrArgAlaHis 41  
QY 145 AAGCGCTCG 195  
Db 42 ProProSerTyrGly-----ArgThrAspLeuSerSerGlyArgLeuTrpVal 58  
QY 196 -----AGTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 225  
Db 59 AsnSerAlaAlaLeuSerProAlaProTyrAlaThrGlyProGlyProAla----- 75  
QY 226 GCGATCCGCGAGCAGCTGTCGAGCGGTGGCGGAGGCGCGGAGCGCGCGCGCGCGCG 285  
Db 76 -----ProSerTyrAlaAlaAlaThrLeuAlaValProGlySerLeuLeuGlyAlaSer 93  
QY 286 GGGCAGCG 327  
Db 94 Gly---GlyLeuAlaGlyAlaAspLeuAlaTrpLeuSerLeuSerGlyGln----- 109  
QY 328 AGCAAGCATATACCG 387  
Db 110 ---GlnGluLeuLeuArgLeuValArgProProTyrSerTyrSerAlaLeuIleAlaMet 128  
QY 388 GGCATCCGCGACTCG 447  
Db 129 AlaIleGlnSerAlaProLeuArgArgLeuThrLeuSerGlnIleTyrGlnTyrValAla 148  
QY 448 GGCAGTTCCTCTTTCG 507  
Db 149 GlyAsnPheProPheTyrLysArgSerLysAlaGlyTrpGlnAsnSerIleArgHisAsn 168  
QY 508 CTTTCGCTCAAGCTGCTTCGCTCAAGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 567  
Db 169 LeuSerLeuAsnAspCysPheLysValProArgAspGluAsnAspPro---GlyLys 187  
QY 568 GACAACTACTGATGCTCAACCCCAACGCGAGTACACCTTCGCGCGCGCGCGCGCG 627  
Db 188 GlyAsnTyrTrpThrLeuAspProAsnCysGluLysMetPheAspAsnGlyAsnPheArg 207  
QY 628 CG 687  
Db 208 ArgLysArgArgArgGlyGlyThrSerGluAlaAlaValProGlyAlaSerSerPro 227



AC Q8K3X9; 4.5e-13 Length: 375  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created) 405.50 Matches: 132  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update) 46.69% Conservative: 37  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) 36.46% Mismatches: 138  
 DE Putative forkhead transcription factor. 55 Indels: 55  
 GN FOXL2. 11 Gaps: 15  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cocquet J.A., Pailhoux E., Jaubert F., Servel N., Xia X.,  
 RA Panetier M., De Baere E., Messiaen L., Cotinot C., Fellous M.,  
 RA Veitia R.A.;  
 RT "Evolution and Expression of FOXL2";  
 RL J. Med. Genet. 0:0-0(2002).  
 DR EMBL; AF522275; AAN04088.1; -;  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001766; TF\_Fork\_head.  
 DR Pfam; PF00250; Fork\_head; 1.  
 DR PRINTS; PRO0053; FORKHEAD.  
 DR PRODOM; PD000425; TF\_Fork\_head; 1.  
 DR SMART; SM00339; FH; 1.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS00039; FORK\_HEAD\_3; 1.  
 DR SEQUENCE 375 AA; 38887 MW; F920C8407686C4B6 CRC64;  
 Alignment Scores:  
 Pred. No.: 4.5e-13 Length: 375  
 Score: 405.50 Matches: 132  
 Percent Similarity: 46.69% Conservative: 37  
 Best Local Similarity: 36.46% Mismatches: 138  
 Query Match: 55 Indels: 55  
 Gaps: 11  
 US-10-087-080-31 (1-1209) x Q8K3X9 (1-375)  
 QY 190 GAACAGAGTGGGAGCGCGCGCGGCGGAGGAGCGGATCCGGGAGCGAGCTGCTGCA 249  
 DB 9 GluAspThrAlaGly-----ThrLeuLeuAlaProGluSerGly 21  
 QY 250 GCGGTGTGGCGAGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 309  
 DB 22 ArgAlaValLysGlu--AlaGluAlaSerProProSerProGlyLysGlyGlyThr 40  
 QY 310 AGCGGCGAGGTGCACGACGACCAATATACGGCGGCGGCGGCGGCGGCGGCGGCG 369  
 DB 41 ThrProGlu-----LysProAspProAlaGln--LysProProGlySerTyr 55  
 QY 370 ATCCGCTCATTCGCATCGCATCCGCGATCGCGGCGGCGGCGGCGGCGGCGGCGGAG 429  
 DB 56 ValAlaLeuIleAlaMetAlaIleArgGluSerAlaGluLysArgLeuThrLeuSerGly 75  
 QY 430 ATCACAGGTACCTCATGGCGAGTCCCTCCCTTTTCGCGGCGGCGGCGGCGGCGGCG 489  
 DB 76 IleTyrGlnTyrIleIleAlaLysPheProPheTyrGluLysAsnLysLysGlyTrpGln 95  
 QY 490 AACTCCGTGCGCCACAACTTTTCCTCAAGACTGCTTCCTCAAGGTGCTGCGCGGAGCCC 549  
 DB 96 AsnSerIleArgHisAsnLeuSerLeuAsnGluCysPheIleLysValProArgGluGly 115  
 QY 550 TCGCGGCGCTGGGCGGACAGCACTACTGATGCTCAACCCACAGCGAGTACACTTC 609  
 DB 116 GlyGlyGlu---ArgLysGlyAsnTyrTrpThrLeuAspProAlaCysGluAspMetPhe 134  
 QY 610 GCCAGCGGGTCTCCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 669  
 DB 135 GluLysGlyAsnTyrArgArgArgArgMetLysArgProPheArgProProAla 154

QY 670 -----CCGGGGCTG-----CGGCGGAGGAGGCGCGGCGCTC 702  
 DB 155 HisPheGlnProGlyLysGlyLeuPheGlySerGlyGlyAlaAlaGlyGlyCysGlyVal 174  
 QY 703 CCG 759  
 DB 175 ProGlyAlaGlyAlaAspGlyTyrGlyTyrLeuAlaProProLysTyrLeuGlnSerGly 194  
 QY 760 GCCCGCCAGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 817  
 DB 195 Phe-LeuAsnAsnSerTrpProLeuProGlnProProSerProMetProTyrAlaSerCy 214  
 QY 818 -----ACAGCATCTCGCAAGCCCTTCGCGAGCGCTCGCTCAGGACACGCGCC 867  
 DB 214 sGlnMetAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlySe 234  
 QY 868 CCGCGGAGGAGCTTCAGTGGGGCG-----CGGCGCGCTCGCGCGCGCTCGCGCG 924  
 DB 234 rProGlyAla-AlaAlaValLysGlyLeuAlaGlyProAlaAlaSer----- 250  
 QY 925 GCGCTCTCCCG 984  
 DB 251 -----TyrGlyProTyrSerArgValGlnSerMetAlaLeuP 263  
 QY 985 GAGCGCGCGCGCTGGCGCGCGGAGCGGAGCGGAGTGCACCGCGCGCGCGCGCTCT 1044  
 DB 263 roProGlyValValAsnSerTyrAsnGlyLeuGly-----GlyProProA 278  
 QY 1045 CTTGCACCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1100  
 DB 278 laAlaProProProProProProProProProProHisProHisProHisAlaHisL 298  
 QY 1101 --CGGCGCGCACTGTACTGCCCGCTCG-----GCTCGCGCGAGCGCTG 1143  
 DB 298 euHisAlaAlaAlaAlaAlaProProAlaProProHisHisGlyAlaAlaAlaProPro 318  
 QY 1144 CAGCGCGCTTAGCCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 1203  
 DB 318 roGlyGlnLeuSerProAlaSer---ProAlaThrAlaAlaProProAlaProAlaPro 337  
 QY 1204 GC 1205  
 DB 337 hr 337  
 RESULT 13  
 QY 093440 PRELIMINARY; PRT; 528 AA.  
 AC 093440;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Winged-helix transcription factor.  
 GN CFKH-1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]\_TaxID=9031;  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=Embryo;  
 RA Buchberger A., Schwarzer M., Brandt T., Pabst O., Seidl K.,  
 RA Arnold H.H.;  
 RT "Chicken winged-helix transcription factor CFKH-1 prefigures axial and  
 RT appendicular skeletal structures during chicken embryogenesis";  
 RL Dev. Dyn. 212:94-103(1998).  
 DR EMBL; Y17794; CAA76851.1; -;  
 DR HSP; G63245; 2HFH.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001766; TF\_Fork\_head.